Detecting Discrete Scale Invariance In Financial Markets

Thesis

Detecting Discrete Scale Invariance in Financial Markets

Christopher Lynch

A thesis presented for the degree of

Doctor of Philosophy

School of Mathematics and Statistics

The Open University

United Kingdom

June 2020

Abstract

This thesis presents methodologies to identify periods in financial markets where the governing regime shows characteristics of discrete scale invariance. Log-periodic power laws often occur as signatures of impending criticality of hierarchical systems in the physical sciences, and it has been proposed that similar signatures may be apparent in the price evolution of financial markets as bubbles form. The features of such market bubbles have been extensively studied over the past twenty–five years, and models derived from an initial discrete scale invariance assumption have been developed and tested against the wealth of financial data with varying degrees of success. In this thesis, the equations that form the basis for the standard log-periodic power law model and its higher extensions are compared to a logistic model derived from the solution of the Schröder equation for the renormalisation group with nonlinear scaling function. Subsequently, a methodology is developed to identify change–points in financial markets where the governing regime shifts from a constant rate-of-return, i.e. from normal growth, to superexponential growth described by a power-law hazard rate. It is suggested that superexponential regimes correspond to financial bubbles and anti-bubbles driven by herding behaviour of market participants. It is from this theory that a predictive algorithm is developed that may have merit in identifying not only when a period of herding behaviour begins, and where it ends. The theory also provides tools which may facilitate profitable trading strategies across a broad spectrum of asset classes.

Dedication

To Laura, Emily, Victoria, and Jemima

Acknowledgements

I would like to thank my PhD supervisors, Dr. Ben Mestel, Dr. and Alvaro Faria, for their help and assistance during the course of my research. It should be noted that Dr. Mestel has collaborated with me on two peer–reviewed publications (Lynch & Mestel 2017 and Lynch & Mestel 2019), the content of which form the basis of Chapters 5 and 6.

I would like to extend a special acknowledgement to Dr. Mestel for the monumental time and effort he has given to my development as a researcher, the enthusiasm he has imparted to me for the subject matter we have pursued together during the course of my thesis, and most importantly the great friendship and support he has shown. And the jokes, of course.

Declaration

Except for where otherwise stated, the results presented in this thesis represent my own work. In the case of any joint work, that which is wholly or mainly my own work is included without further comment. None of this research has previously been submitted for a degree or other qualification to this or any other university or institution. Material from Chapter 5 of this thesis has been published as joint work in the peer-reviewed articles Lynch, C. & Mestel, B. (2017). Logistic Model For Stock Market Bubbles And Anti-Bubbles, *International Journal of Theoretical and Applied Finance* 20(6), and material from Chapter 6 has been published as joint work in the peer-reviewed article Lynch, C. & Mestel, B. (2019). Change-point analysis of asset price bubbles with power-law hazard function, *International Journal of Theoretical and Applied Finance* 22(7).

Contents

Chapter 1

Introduction

In 1996, two independent research groups published papers (Sornette et al. 1996 and Feigenbaum & Freund 1996) which focused on the evolution of the Standard and Poor's 500 index (the **S&P500**) up to the global financial crash of October 1987 ("Black Monday"), and noted that the long-term price development prior to the crash could be described phenomenologically by a power law modulated with log-periodic oscillations. It was proposed that this suggested that a crash might be viewed as a critical point in the behaviour of a hierarchical network of highly interconnected agents, and that the price bubble formation prior to the crash was explained by the network becoming self organised, or exhibiting some form of cooperative or herding behaviour between the agents.

The genesis of this idea came from research in the field of statistical physics, and in particular from research into the renormalisation group equations to describe transitions of theoretical models. It was expected that these models might describe the many physical systems that undergo rapid or instantaneous transitions when an influencing factor reaches some critical value, for instance, the critical temperature below which certain metals become spontaneously magnetic, or the stress leading to catastrophic failure or rupture of materials.

The thrust of this research was the behaviour of the system around these critical points, and the proposition that to understand this behaviour one should not look at the system as a series of consequential events leading to the critical event, i.e. one rupture or failure in a material causing a rupture in its locality

and so on, but rather as a global framework driven by cooperative behaviour of elements acting in a self-similar way over many discrete scales, i.e. microscopic effects across the whole system culminating in the critical event. It is this so-called discrete scale invariance (**DSI**) which necessarily leads to the log-periodic oscillatory phenomena around critical points. The theoretical basis of this idea is discussed in detail in Chapter 3.

It has been useful in the development of the discrete renormalisation group theory of critical points in relation to both natural systems and financial markets to consider the two-dimensional square lattice Ising model for magnetic phase transition as a proxy for reality in these circumstances, since it is a simple statistical model that exhibits phase transitions and self-similarity over a discrete range of scales. In his 1924 thesis (Ising 1924), Ising proved that the one-dimensional lattice model does not undergo any phase transitions and assumed that this was the case for models of higher order. However, in 1944, Lars Onsager (Onsager 1944) found an analytic solution for the two-dimensional model and showed that phase transitions were possible in this case. As discussed later in §3.2, the theory underlying the log-periodic power law model for financial bubbles was based on the diamond lattice Potts model, which is a generalised form of the Ising model.

Laboratory studies into certain characteristics (such as acoustic emissions) of materials under stress prior to failure led to hypotheses of self-similarity in the field of seismology. It was suggested (Newman et al. 1993 & Newman et al. 1995) that plate tectonics under stress could also be view as a hierarchical system with long-range effects contributing to critical behaviour. In support of this theory, log-periodic oscillations were detected in characteristics such as chlorine concentration levels in water systems in the vicinity of the epicentre of the Kobe earthquake of 1995 (Johansen et al. 1996), an account of which is given in §2.1. Further examples of physical occurrences of log-periodic power laws can be found in Wilk & Wlodarczyk (2017) and Smolla et al. (2020).

It was this research into log-periodic behaviour in natural systems that, in 1996, led researchers to extend this idea of DSI in highly interconnected systems to applications in financial markets, when it was noticed that the long bull run

of the S&P500 prior to the Black Monday crash of 1987 clearly followed the faster than exponential growth typified by a power law moderated by log-periodic fluctuations, or, as it is referred to in this thesis, a log-periodic power law (**LPPL**).

Ever since the global stock-market crash of 1987, commentators have not been able to agree on the fundamental reason for this large correction in the market. Some proponents of the theory of discrete scale invariance in financial markets postulated that perhaps financial bubbles could be described by herding behaviour of market participants leading to self-similarity of the system at the boundary of a phase transition, i.e. a financial crash. The point was that there was no reason for the crash other than that "stress" had built up in the system during the formation of the bubble. This bubble led to instability so that, at the point of the critical event, all agents in the market were aligned, taking only one seller to bring the market down – the straw that breaks the camel's back.

Following the initial 1996 work, the huge amount of data available across many financial markets has led to an explosion of research in this field resulting in strongly suggestive but inconclusive evidence of DSI in financial markets, for, despite this ocean of available data, the occurrence of bubble and crash regimes are actually quite rare. Therefore, after two decades of research the question that remains to be answered is whether these patterns arise merely by chance, or is there any significance of the observations that have been made? Furthermore, what other models are implied by the renormalisation group formalism and can these one uncover deeper signals of self-similarity in financial markets that would allow prediction of future events? As will be seen in the following chapters, there might be tantalising evidence of this possibility.

This thesis first outlines the history of financial crashes as critical points from its origins in statistical physics in the context of the natural sciences, and how these ideas have courted controversy and heated debate over the past quarter of a century (Chapter 2). Next, in Chapter 3, the theoretical foundations are laid down in summary form. This is followed by a presentation of methods to fit LPPLs to financial time series in Chapter 4, before the research on which this thesis is based is described.

The results of this research are presented in Chapters 5 through to Chapter 8, which includes a presentation of a new LPPL model derived from a non-linear perturbation to the commonly assumed linear form of the renormalisation group equations. The classic LPPL equation and its higher extensions are compared to this model which has been derived from the solution of the Schröder equation for the renormalisation group with nonlinear scaling function. Results for the S&P500 and Nikkei 225 indices studied previously in the literature are presented in Chapter 5 and compared to established models, including a discussion of the apparent frequency shifting observed in the S&P500 index in the 1980s. In the particular case of the Nikkei 225 anti-bubble between 1990 and 2003, the logistic model appears to provide a better description of the large-scale observed features over the whole 13-year period, particularly near the end of the anti-bubble.

This is followed in Chapter 6 by an investigation into methodologies to determine whether financial markets are in an imitative governing regime and thus exhibiting DSI, concluding with a novel approach based on maximising the log-likelihood parametrisation of the log-return of DSI exhibiting markets modelled by a power-law hazard function. This methodology is used to identify change-points in financial markets where the governing regime shifts from a constant rate-of-return, i.e. normal growth, to superexponential growth described by a power-law hazard rate, the latter regime corresponding to financial bubbles driven by herding behaviour of market participants. Assuming that the time series of log-price returns of a financial index can be modelled by arithmetic Brownian motion, with an additional jump process with power-law hazard function to approximate the superexponential growth, a threshold value of the hazard-function control parameter is derived, allowing identification of the most likely market regime at any given time. An analysis of the S&P500 index over the last 60 years provides evidence that the methodology has merit in identifying when a period of herding behaviour begins, and, perhaps more importantly, when it ends.

The research concludes in Chapter 7 with a presentation of an approach to predicting future market movements. This is based on the likelihood that the prevailing governing regime is exhibiting DSI as determined by the previously

described power-law hazard function detection method. In this method, a computationally efficient proxy to the theoretical model is derived, enabling the development of a predictive algorithm that can be applied in real-time scenarios. Back-tested results across a wide range of financial markets are presented in the Chapter 8, providing a tantalising glimpse at what might be achieved as this theory is developed.

The final chapter concludes this thesis and comments on the merits of the research in providing a methodology to detect discrete scale invariance, and to predict future price evolution of financial markets. The thesis is brought to a close with some suggestions for further research.

Chapter 2

Historical context

2.1 Discrete scale invariance and criticality in natural systems

The renormalisation group method to study phase transitions in systems that exhibit critical phenomena has been available for many decades, and seeks to describe such systems in terms of scaling factors, i.e. the large scale macroscopic behaviour is derived by "averaging" the microscopic effects over many scales. It had been found early on in the development of these methods that the renormalisation group equations for a thermodynamic property, denoted say, by E, of a system in terms of a dependence on temperature, T, imply singular behaviour around a critical point, T_c, in the form of a power law (Nauenberg 1975) such as

$$E(T) \propto |T - T_c|^{\alpha} \tag{2.1}$$

Nauenberg (1975) gives extensive examples of studies that have shown this result both theoretically and experimentally and, in particular, cites Wilson (1971) for showing dimension-dependent universality of these critical exponents. Importantly, Nauenberg (1975), along with Jona-Lasinio (1975), noticed that the system can be shown to have log-periodic dependencies in $|T - T_c|$ by allowing α to take complex values rather than the previously accepted assumption of $\alpha \in \mathbb{R}$. Nauenberg (1975) briefly discusses former experimental results in physi-

cal systems exhibiting critical phenomena that have shown this scaling with real valued critical exponents, but it took the work of Anifrani et al. (1995) to provide evidence of the expected log-periodic oscillations in real-world physical systems.

Anifrani et al. (1995) focused on log-periodic acoustic emissions emanating from carbon-fibre spherical tanks placed under stress, and how the development of these emissions could be used to predict rupture. They conjectured that rupture is not a single event, but an accumulation of small effects, the results of which can be predicted only by looking at the system as a "global framework" being governed by cooperative behaviour, rather than a succession of causally connected events. This implies that fractures can form across the material at randomly distributed pre-existing flaws. It is not the growth of one of these fractures that cause the critical failure but rather the interaction of smaller fractures over a range of scales that lead to a critical density of fractures in the locality of the failure point. By taking a renormalisation group approach to this problem, the authors were able to confirm the existence of log-periodic corrections to the acoustic emissions in the non-critical (i.e. before rupture) region. The resulting models could predict the failure pressure within a 5% error, and, interestingly, when the pressure had reached only 85% of the observed failure pressure. Their results were claimed to be the first laboratory observations in a natural context of such discrete scale invariance, and perhaps led the way for investigations into other physical systems, most notably in the field of seismology.

Sornette & Sammis (1995) followed this theme and investigated the nature of ruptures associated with earthquakes. At this time, several authors had theorised that earthquakes could be viewed as hierarchical systems displaying critical phenomena which could be described by renormalisation group models, of which Newman et al. (1993) provides a good summary. Furthermore, an account of the empirical evidence of universal scaling laws in seismology (such as the power-law distribution of earthquake aftershocks and the fractal properties in the geometry of earthquakes faults) can be found in Kagan (1991).

Building on this previous research into earthquakes as critical points, Sornette & Sammis (1995) looked at the power-law fit by Bufe & Varnes (1993) of the

Benioff strain, ϵ, a measure of the stretching of the Earth's surface, to the 1989 Loma Prieta earthquake and to other seismic activity in the Alaska-Aleutian region. This power law is given by

$$\epsilon = A + B|t_f - t|^m \tag{2.2}$$

where A, B and m are constants and t_f is the time of failure, i.e. the earthquake.

The authors point out the Bufe's power law was not itself predicated on an assumption of scaling near a critical point. However, they recognised that this behaviour was indicative of a preferred scaling, and so a better fit of the data could be achieved by inclusion of the log-periodic corrections to such scaling. It is in this paper one finds the first instance of the log-periodic power law derivation which pervades the entirety of the subsequent twenty years of research and is the theoretical foundation of this thesis.

Sornette & Sammis (1995) state their log-periodic power-law model (see Section 3.3) for Benioff strains in terms of three linear parameters A, B, and C, and a further four nonlinear parameters m, the critical exponent, $\lambda = 2\pi/\omega$ (where ω is the angular log-frequency) and finally the phase angle, ψ, as follows

$$\epsilon = A + B(t_f - t)^m \left(1 + C\cos\left(\frac{2\pi}{\log\lambda}\log\left(t_f - t\right) + \psi\right)\right) \tag{2.3}$$

This is the fundamental form on which all subsequent models were built. On presenting this equation the authors asserted that if their seven-parameter model provides better predictive performance of the time of failure (the earthquake) than the four parameter model of the simple power-law, then they must conclude that the information contained in their higher dimensional model is significant. One could argue that, at this early stage of research and with such a small data set, this assertion is doubtful. It is perhaps interesting to note that this confidence followed Sornette into his research concerning bubble formation in financial markets and there have been strong challenges to similar statements regarding the significance (statistical or otherwise) of his and his co-workers' results.

Nevertheless, the authors produced positive results based on Bufe's data sets

in that the value of the time to failure, i.e. the time of the earthquake given by the best fit (using the Levenberg-Marquardt algorithm) more accurately matched reality than did the best fit of the power-law model. It should be noted here that there is no discussion in this paper of what constitutes a "best fit" of such multi-dimensional, nonlinear models, and, as such, it is unclear what information the value of any parameter really conveys when there are so many possible similarly valued local minima across the parameter space.

Saleur et al. (1996) followed soon after the publication of Sornette & Sammis (1995) and expanded on, and more fully explained, the theoretical framework on which the latter work was based, drilling down into the detail of how discrete scale invariance leads to a complex fractal dimension implying log-periodic corrections close to the critical point, and explaining in detail the philosophy behind the renormalisation group formalism in this context. Apart from this, the paper put the whole theory of earthquakes as critical points on a much firmer footing by considering the underlying physical interpretation of the parameters and addressing some of the concerns a sceptical reader might have regarding the significance of the better model fits with such an increase in the number of parameters[1].

Saleur et al. (1996) laid the groundwork for the research to follow on criticality in financial markets, so it is worthwhile spending a little time understanding the authors' fundamental ideas, as they argue that it is the firm theoretical basis on which the model is founded which should alleviate concerns that the goodness of fit is a by-product of the size of the parameter space.

It has already been noted (Sornette & Sammis 1995, Anifrani et al. 1995) that in material science there is evidence that correlations of activity exist prior to critical events over scales ranging from the very small to the very large. Saleur et al. note that similar long range correlations of increasing magnitude seem to precede a subsequent major earthquake. The underlying assumption is that "each earthquake represents failure of some interior region the size of which scales with the size of the event", and that "the failure of its local region is a part of the precursory failure sequence of an even larger event". As remarked before, this is

[1]One may be reminded of a quote attributed to John von Neumann by Enrico Fermi "With four parameters I can fit an elephant, and with five I can make him wiggle his trunk."

quite different from a power-law model derived from an assumption of sequential casually related events (Anifrani et al. 1995), with the fundamental idea being that the renormalisation group formalism provides more information in the form of "universal" recognisable signatures than would a power-law only.

This interaction of a highly interconnected network of regions suggested to the authors that such a geometry could be modelled by considering the exact solution to simple "spin models" such as the diamond-lattice Potts model (see §3.2). In finding the exact solution, the functional form of the renormalisation group equations are recovered and after some careful argument, it is concluded that this supports the original hypothesis of the applicability to earthquakes of the renormalisation group formalism.

Equally as important to geometric discrete scale invariance in these networks is the discrete scale invariance in the "critical dynamics" coupled to this geometry. When one talks about these critical dynamics one is thinking about the way in which a fault propagates and the way these dynamics scale rather than the scaling of the fault's particular geometry. The authors demonstrated that it is possible to generate log-periodic behaviour without an underlying complex dimensioned fractal geometry but as a consequence of random network dynamics. The important point here is that there is no expectation that the system is required to have discrete scale invariance built into its geometric fabric, but DSI can arise spontaneously from the inherent disorder in the system.

Saleur et al. (1996) again use the example of the Loma Prieta earthquake to illustrate their results, and reiterate the superior nature of their data fit and forecast of the impending earthquake, and here accept that one might expect a better fit could be achieved with such a larger choice of parameters. However, they assert that the theoretical basis of the model and the clear geometric series of foreshocks indicate the validity of the model. Furthermore, they do not present the model as a predictor of earthquakes, but rather expect log-periodic behaviour to be present in earthquake data.

Further support for the prevalence of log-periodic fluctuations in earthquake dynamics came on the publication of Johansen et al. (1996) which made a study

into ground water ion concentration close to the epicentre of the Kobe earthquake in January 1995. These changes in hydrogeochemisty and other non-seismic phenomena such as permeability enhancement and groundwater discharge fluctuations have been known to exhibit critical behaviour close to a seismic event.

The experiment studied high-quality mineral water produced in the mountains near to Kobe that had been collected and bottled from 18 months before the earthquake. The ion concentration level in these bottled waters showed an acceleration decorated by log-periodic fluctuations of concentration similar to that of the Benioff strain data for the same period. As has already been seen in the literature, the authors again stress the point that, from statistical physics, one expects to see power-law acceleration of observations up to the critical time due to similarity over many scales, but is it only in hierarchical systems that one sees log-periodicity of observations derived from discrete scale invariance.

Fitting (2.3) to the ion-concentration data was achieved by minimising the root-mean-square error analytically by treating the constants A, B, and C, as functions of the other nonlinear parameters. This has become standard practice when considering these log-periodic models, as has the taboo search of the nonlinear parameter space for local minima, and the use of the nonlinear least-squares Levenberg-Marquardt algorithm. By choosing (what the authors considered to be) the global minimum, they were surprised to find the predicted value of t_c to be only two days later than the actual date of the earthquake. Testing against the null hypothesis of the constant but noisy chlorine content, and an alternative null hypothesis of exponential growth, the authors concluded that these hypothesises could be rejected, and against the results of other tests performed, such as the bootstrap re-sampling scheme, it was concluded that the claim of log-periodic oscillations in the data could be supported.

In their conclusion, the authors remark that the proof of any predictive power of this model is in not in its ability to fit data closely, as would be expected where the parameter space is large, by rather to find good fits with some universality in the parameter values.

However, effective earthquake prediction methodology remained elusive, maybe

due to the lack of empirical data, and many of the proponents of the development of this theory turned their attention to applications in the financial markets, where there is an abundance of data sets to study and a stark example of log-periodic behaviour in the long bull run up to the great crash of 1987, Black Monday.

2.2 Log-periodic power laws and financial markets

The work of Sornette et al. (1996) and Feigenbaum & Freund (1996), which spawned the ubiquitous log-periodic power-law (**LPPL**) models for bubble formation in financial markets, had their origins in phenomenological observations of the S&P500 index in the period leading up to the crash of October 1987 and a similar time series progression in the Dow Jones index preceding the great crash of 1929.

Using the renormalisation group theory of criticality in hierarchical systems that had been extensively studied in natural systems, the authors of these works applied a discrete scale invariance assumption of the underlying mechanisms driving financial markets to derive a model describing the periodic fluctuations around a faster than exponential development in prices observed in financial time series prior to many of the significant market collapses. The log-periodic signatures observed in the two time-series in question suggested to the authors there were mechanisms underpinning these markets that one could hypothesise should be hierarchical in nature implying a complex fractal dimension, i.e. discrete scale invariance.

Assuming a constant-return hypothesis one expects to see exponential price development, but during periods in which the market price is in an accelerating regime prior to a large drawdown, price development was observed to follow a power law. It was suggested in Sornette et al. (1996) that this unsustainable growth up until some critical time is a consequence of the tendency for traders or market professionals to exhibit imitative or cooperative behaviour, and it is

this mechanism which drives the hierarchical nature of financial markets and the observed log-periodic oscillations are a consequence. This fundamental expectation has been the genesis of the last twenty years of discussion on the matter of financial crashes as critical points.

Both papers consider the log-periodic oscillations prior to the Black Monday crash of October 1987 as it effected the S&P500 index, and achieve similar results when fitting equation (2.3) to the observed data. At the time, this crash is singled out as being somewhat exceptional since, although the US stock market fell by 30%, stock prices across the globe also suffered major drawdowns despite the diverse nature of local stock markets. It is a matter of debate as to what was the root cause of the crash, but Sornette et al. (1996) cite statistical evidence which indicates an aggregated measure of monthly returns on global stock markets was highly correlated to each local market movements since the beginning of 1981 to the month before the 1987 crash. Perhaps this, the authors note, suggests a worldwide co-operative phenomenon driving financial markets leading to the same kind of precursory signatures that are found in the natural world.

Feigenbaum & Freund (1996) also consider the S&P500 index for the 1962 crash and the Dow-Jones index leading up to the great crash of 1929. As a comparison, the authors of this paper also look at the more benign period between 1990 and 1994 on the S&P500 index. In contrast to the previous paper, the authors do not dwell too much on hypothesising about the underlying market mechanisms that could bring about these observations, but pragmatically note that these signals certainly are present prior to these crash events, and cannot be detected in the single period they chose where no crash event happen then or soon after.

Notwithstanding their similar results, there is a fundamental assumption in Sornette et al. (1996) which is claimed to be unfounded in Feigenbaum & Freund (1996). In the former, the hierarchy of characteristics is expected to lead to universality in the physical parameters, in particular the angular frequency $\omega = 2\pi/\lambda$. However, the latter show that this is not true in the case of the S&P500 index since the "best fit" for equation (2.3) up to two years prior to the crash is

very different from the fit using data going back eight years, if one decides that the critical time t_c should be close to that of the time of the crash. This leads to a "over-oscillation" of the model in the later years if using a longer term fit.

Sornette & Johansen (1997) build on some of the ideas suggested by Sornette et al. (1996) regarding the mechanisms that could lead to discrete scale invariance in financial time-series and extend the original LPPL model (referred to here as the first-order model) to take account of the apparent frequency shift in the log-periodic oscillations observed in the S&P500 index from the beginning of 1980 to the crash of 1987.

The authors' hope in this paper was that, in examining the financial markets in their most stressed state, i.e. that of an unstable equilibrium prior to a crash, information about their underlying structure and organisation would be revealed to support their theory that crashes have their origins in the herding, or "collective organisation" behaviour of traders.

The authors point out that Feigenbaum & Freund (1996) had found that the log-frequency of the log-periodic oscillations tends to decrease as the crash is approached. To account for the apparent frequency shift, Sornette & Johansen (1997) developed a nonlinear correction to the first-order model motivated by Landau expansions from the theory of phase transitions in statistical mechanics. For example, by taking the derivative of the power law $F(x) = Ax^{\alpha}$ with respect to $\log x$ gives $dF(x)/d\log x = \alpha F(x)$. Clearly, for real $\alpha > 0$, $|F(x)|$ increases with $\log x$, and conversely when $\alpha < 0$, $|F(x)|$ decreases with $\log x$. Therefore, $F(x) = 0$ is an unstable equilibrium when α is positive, and a stable equilibrium when α is negative. Therefore, regarding α as a bifurcation parameter, the point $\alpha = 0$ is a bifurcation point for a saddle-node bifurcation. Unfolding of the bifurcation to order $O(|F|^3)$, leads to the following model equation:

$$I(t) = A + B\frac{(t-t_c)^{\beta}}{\sqrt{1+\left(\frac{t-t_c}{\Delta_t}\right)^{2\beta}}}\left(1 + C\cos\left(\theta(t)+\phi\right)\right), \text{ where}$$

$$\theta(t) = \omega\log\left(t-t_c\right) + \frac{\Delta_\omega}{2\beta}\log\left(1+\left(\frac{t-t_c}{\Delta_t}\right)^{2\beta}\right), \qquad (2.4)$$

This equation, referred to as the second-order LPPL model, is derived fully in Section 3.3. One can see that as the critical time is approached, the log-frequency of the oscillations shift from $\omega + \Delta_\omega \to \omega$.

Further unfolding of the bifurcation to order $O(|F|^5)$ gives the so called third-order model and this is derived by Johansen & Sornette (1999b). However, this model is rarely used for fitting observed data elsewhere in the literature. Johansen & Sornette (2000) remark that on 25th January 1999, they publicly predicted a trend reversal in the Nikkei 225 index based on the best fit of the third-order model over the 8 year period between 1990 and 1998. Extending the fit past 1998 into 1999 showed quite remarkable performance, and Johansen & Sornette (2000) claimed their results were convincing when looked at from the perspective of a neutral Bayesian observer. However, as Lynch & Mestel (2017) show, this predictive nature does not continue for the third-order fit, and much better results are possible with the logistic model. In this thesis, the results of Lynch & Mestel (2017) are presented in more detail. Matsushita et al. (2006) also use use a variation of the third order model (in this case the third harmonic of the Taylor expansion, which is analogous to model derived from the Landau expansion but does not retain the physical properties of the parameter space) to fit an apparent bubble formation in the BRL/USD exchange rate between 2000 and 2005.

In fitting (2.4) to the data (Johansen & Sornette 2000), the constant average growth over the period was filtered out by taking the logarithm of the index, as it was argued that this was not a phenomenon that they were trying to detect. Furthermore, the authors made important restrictions to the acceptable parameter ranges. Firstly, they rejected large values of ω since this would fit oscillations at very small scales, and secondly, acceptable range for the critical exponent was restricted to $[0, 1]$ since in this range the index will remain bounded, which is the behaviour observed in the real world, leaving the possibility for a singularity in its derivatives. The result of the study was that the second-order model did indeed perform better than the first order model. However, as can be seen in Lynch & Mestel (2017), this observed frequency shifting is not necessarily a phenomenon that has any basis in reality, but perhaps merely a result of the linearisation of

$\phi(x)$ in the first-order model.

In their conclusion the authors restate their previous theory regarding the mechanism leading to a crash to be the co-operative behaviour of market participants, but expand on this idea and suggest a "weak efficient market hypothesis" in that the individual ignorance of market traders does indeed lead to a collective macro intelligence, and that the information of impending crashes is contained in the market dynamic and can be observed over very long time scales, in contrast to very short time scales used in previously proposed market crash models. Furthermore, they theorise that market dynamics are purely endogenous, and exogenous factors act as mere triggers. However, the authors do not present much more evidence to support their theory, but point to close observed values for the critical exponent in both the 1987 and 1929 crash as evidence which defends their position.

Shortly after this paper was published, the other original authors produced a brief note (Feigenbaum & Freund 1998) describing similar log-periodic signatures apparent in the S&P500 index correction of October 1997. Interestingly, this much less cited article, looks at a saw-tooth rather than an harmonic regime of log-periodic activity. The authors filter a power law from the observed data and estimate the periodicity of the resulting function in terms of $\log(t_c - t)$ and find similar angular log-frequencies as they found in their previous studies of the 1929 and 1987 crashes. It transpired that claims by Sornette et al. (along with several others) that this crash could have been predicted led to some heated debate.

The following couple of years, 1998 and 1999, saw a flurry of papers being published covering a wide range of related ideas in the study of financial market crashes as critical points in self-similar systems, along with evidence presented of discrete scale invariance signals being present at much shorter ranges than have previously been considered. This phenomenon was presented by Drozdz et al. (1999) who provided evidence that less dramatic intermediate drawdowns showed signs of log-periodic oscillations when analysing the German stock market in 1998. It is interesting to note that the authors detected "universal" preferred scaling ratios in this data. It is recognised that there is no contemporary agree-

ment as to what actually constitutes a crash, and perhaps more importantly, the lack of research into how features of a particular crash can be absorbed into the description of a larger event some time later. To overcome these issues, it is suggested that market dynamics are in a constant regime of boom or bust, but on varying scales in time and magnitude, and, in fact, this suggests a more general nonlinear renormalisation group flow map could be derived rather than the linear λx that from which the log-periodic equations had up until then been derived. Lynch & Mestel (2017) present a nonlinear perturbation to the linear renormalisation group flow map, as is explained in Section 4.4. In common with Lynch & Mestel (2017) the authors pick out the gross features of the time-series by eye and use sequential peaks or troughs to estimate a log-angular periodicity and, in the analysis of their data, conclude that there is little qualitative difference between the progression of sequential peaks and troughs and the peaks and troughs that are intermediate, and indeed present similar log-angular periodicity. One issue that can be thought of here is that there is no systematic way of deciding which peak or trough is used to create the geometric progression.

Variations on the theme of discrete scale invariance in financial markets which allude to the predictive features of these methods, were published by most notably by Gluzman and Yukalov (Gluzman & Yukalov 1998b and Gluzman & Yukalov 1998a), along with a series of papers by Vandewalle et al. (Vandewalle, Ausloos, Boveroux & Minguet 1998, Vandewalle, Boveroux, Minguet & Ausloos 1998 and Vandewalle et al. 1999). Gluzman and Yukalov consider using self-similar approximation theory from theoretical physics to analyse time series, and examine how these methods may be applied, very much in a predictive sense to the stock market crashes of October 1987 and 1997, particularly opting for these non-stationary progressions due to the unsuitability of stochastic methods in these circumstances. "Resummation" in this context is simply a variety of the renormalisation group method in which the terms of a divergent series of functions are scaled to another convergent functional series. This series can be mapped to the original function under an integral transformation. In practical usage, the authors take between three and six data points and apply the method

across a wide variety of market situations around the globe, in each case testing whether or not their approximation of the final evolutionary step in the particular market agrees with what happened in reality. In all except one case the error is within 5%, from which they conclude that resummation techniques may be used in analysing financial markets.

Vandewalle, Boveroux, Minguet & Ausloos (1998) suggest stock market dynamics should be viewed as analogous to specific-heat phase transitions. It would appear that the driver behind this is the stark similarity between the "shape" of specific heat capacity evolution close to a phase transition and that of the Dow Jones Industrial Average from 1980 to the crash of October 1987. It should be noted that in the $d = 2$ Ising model, the critical exponent, α, is zero, so that the model follows logarithmic divergence rather than a power-law. It is this logarithmic divergence that is the focus of the analysis that is presented by Vandewalle, Ausloos, Boveroux & Minguet (1998). The authors defend this choice of logarithmic over a power law evolution of the bubble regime as more plausible from analogy with physical systems, but also note that not only does this reduce the parameter space, but builds in universality in the critical exponent. This theory fits in quite well with results of the Nikkei anti-bubble between 1990–2003 when analysed with the logistic model of Lynch & Mestel (2017).

The end of the initial phase of research into this topic is was marked by a paper by Johansen & Sornette (1999c) in which they gave a good summary of what was then the state of the art in LPPL modelling of financial markets, and gave guidelines on using this theory in practical applications. Firstly, the paper refers to the statistical evidence presented by Johansen & Sornette (1998) showing that stock market crashes are outliers and can be thought of as operating in a totally separate regime to "normal" market dynamics. With this in mind the authors go on to lay the important groundwork that models should to encompass both the rationality of market participants, i.e. traders are not a blind herd, but act with intelligence given the available information, and stress the importance of time reversal symmetry-breaking, in that bubbles tend to build up slowly and crash very rapidly, and one must accept that there is a definite arrow of time

which is irreversible. To this they added the first mention of how the interaction of traders imitating their local connections builds up the likelihood of a crash occurring, and this likelihood is referred to as the crash *hazard* rate. The novel point here is that the critical time t_c was no longer considered as the time of the crash, but as the most likely time the crash should occur given that it has not already done so. It is perfectly reasonable to assume that the bubble is not marked by a crash, but this end is the most probable time for such an event to happen.

Johansen & Sornette (1999c) recognised that it was inevitable that probabilistic models were required where agents act on a flow of available information which is stochastic in nature, and, as such, the subsequent iteration in the evolution of the topic was a more extensive and rigorously presented theory which became know as the Johansen-Ledoit-Sornette (JLS) model of rational expectation bubbles with finite-time singular crash hazard rates.

2.3 Johansen-Ledoit-Sornette Bubble Model

The evolution from LPPL models of stock market bubbles into the JLS model began with the rational expectations approach of Johansen & Sornette (1999a). It is argued that, since most draw downs can be modelled by existing stochastic methods, it is reasonable to expect different modelling techniques to be required for the outlying (or extraordinary) draw down of a "crash" and the associated precursory signatures. The technique they describe has the features that underpin the LPPL models, i.e. self-similarity over a preferred range of scales derived by local herding behaviour of market participants, but adds the notion of a crash hazard rate, $h(t)$, "the probability per unit time that the crash will happen in the next instant if it has not happened yet". If the probability that a crash has not occurred up until time t is given by the function $P(t)$, then the hazard rate is expressed as

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t) - P(t + \Delta t)}{\Delta t P(t)} \tag{2.5}$$

This idea is weaved into a differential equation describing the asset price, $p(t)$ dynamics

$$dp = \mu(t)p(t)dt - \kappa\left[p(t) - p_1\right]dj \tag{2.6}$$

where $\mu(t)$ is a function driving price returns, and κ is an assumed fixed percentage of the difference between the price at time t and a reference price, p_1, the amount by which the asset is expected fall if a crash occurs. Here the function j is zero before a crash and one after. However, we can also say that as the risk of a crash increases, traders will expect higher return, and additionally that the martingale condition, where given all the available information up until time t, the expected price at $t + \Delta t$ is $p(t)$, such that

$$u(t)p(t) = \kappa\left[p(t_0) - p_1\right]h(t) \tag{2.7}$$

where t_0 is some initial condition. Combining (2.6) and (2.7) then integrating gives

$$p(t) \approx p(t_0) + \kappa\left[p(t_0) - p_1\right]\int_{t_0}^{t} h(s)ds \tag{2.8}$$

The instantaneous likelihood, $h(t)$ of a crash occurring accounts for the inherent random behaviour of traders, and it is when this imitative behaviour and randomness is in balance that the markets behave in their usual manner. However, when order is found, imitation becomes pervasive and all traders "agreeing" to sell triggers the crash.

The authors proposed that this hazard rate is governed by the number of "connections" a typical trader has such that $dh/dt = Ch(t)^{\delta}$ with $\delta > 1$. Solving this differential equation gives

$$h(t) = \frac{B}{(t_c - t)^{\alpha}}, \quad \alpha = \frac{1}{\delta - 1} > 0 \tag{2.9}$$

where as t tends to the critical time, t_c, $h(t) \to \infty$. Here, there is a shift in methodology from describing the stock *price* development as being driven by the microscopic behaviour of traders, described by the renormalisation group equations, to a macroscopic power law expression of the hazard rate governed by

a measure of a typical trader's interactions. Now t_c is described as not being the actual time of the crash, but the time when the crash is most likely to happen. This is important as it means there is a finite probability that the bubble ends without a crash, i.e. for some initial tine, t_0

$$1 - \int_{t_0}^{t_c} h(t)dt > 0 \tag{2.10}$$

and given sufficiently large returns, a trader may consider it is rational to enter into, or hold a long position as the probability of a crash occurring increases.

Referring back (2.8) it can be seen that on integration of h the exponent becomes $1 - \alpha = \beta$. In order for the price, $p(t)$, to remain bounded with a singularity at $t = t_c$ in its derivative, $\beta \in (0, 1)$.

Johansen & Sornette (1999a) add log-periodicity modulation to the power law hazard rate by loosely arguing complex critical exponents as a feature of hierarchical systems, and by incorporating this into the dynamical equation to recover the first-order equation on integration.

Johansen et al. (2000) supply more detail on the ideas presented by Johansen & Sornette (1999a), and the model so presented became known as the Johansen Ledoit Sornette model of rational expectation bubbles with finite-time singular crash hazard rates. Johansen et al. (2000) reiterated the point that given an analysis of draw downs in financial markets, one should expect events such as the crash of 1987 to happen perhaps once every 180 years, whereas, at the time of writing the US markets had had three in under a century. Again this suggests normal markets are governed by a separate process to those that govern crashes of this magnitude. To test this hypothesis, the authors simulated 10,000 GARCH(1,1) model instances with Student's-t noise over 100-year time periods, and only twice were there instances of three draw downs of "crash" proportions. However, these draw-downs were seen to be fairly symmetric, unlike observed crashes, and moreover, no log-periodic signatures were detected.

From this point onwards, there was little development in the underlying model, and attention shifted to detecting signals of log-periodicity and discrete

scale invariance in the various instances of financial markets experiencing a bubble followed by a crash, or indeed, an "anti-bubble". Sornette & Zhou (2002) analysed the similarities between the S&P500 index between 1996 and 2002, and the Nikkei 225 index between 1985 and 1992, presenting evidence of "herding" in the S&P500 index since August 2000. In pursuing this analysis, they employed several techniques in addition to the usual first and second-order LPPL fits, namely analysis of the geometric progression of subjectively chosen peak and troughs as used by Drozdz et al. (1999) and Lynch & Mestel (2017), a log-spectral analysis of the data de-trended by a pure power-law using a Lomb periodogram, and an interesting spectral analysis of the so called (H, q)-derivative, which is given by

$$D_q^H f(x) \triangleq \frac{f(x) - f(qx)}{((1-q)\,x)^H}, \quad x = t_c - t \tag{2.11}$$

By scanning values of H and q, for various values of t_c, the authors could detect which values of t_c gave the most consistent primary log-frequency over the range of values for H and q. This would seem to be a non-parametric detrending of the data.

Finally, the authors consider the second harmonics in both periodograms of the parametric and non-parametric detrendings. It appears that in both cases, if the log-frequency is determined as ω then there is a second harmonic at 2ω, and an additional term to the LPPL model. The revised model gives better r.m.s. residuals and a Wilks test determines that the null hypothesis of the additional term begin equal to zero should be rejected (Zhou & Sornette (2005) cast doubt on the validity of this test). Using the similarities between the two indices, it was predicted that the US market had not, at that time, entered its frequency shifting phase, as the Japanese market had done so 2.5 years into the anti-bubble, and they propose a sense of the direction of the US market. It transpired their prediction was completely wrong and the market went into a long rally until 2007. Zhou & Sornette (2003a) continue with the non-parametric analysis of the (H, q)-derivative as applied to seven historical bubble-crash regimes and find a universal preferred scaling across these events which agrees well with the LPPL

model fits. The authors also use a Hilbert transform of the data to perform a similar analysis.

Zhou & Sornette (2003b) continue where they left off, but add more focus to the five "crash" like spikes in the S&P500 index anti-bubble between 2000 and 2002. Since their current models (including adding the additional harmonics) could not account for these features, which one can also see in the Nikkei bear run between 1990 and 2003, they propose extensions to the first order LPPL model to account for these intermediate critical points. Up until this point, the underlying theory only accounts for the market dynamics up to the critical point, and is silent on how the market develops after this singularity. The authors argue that their present approach is more powerful since it "shows a deep connection between the five crashes on the one hand and the overall log-periodic power-law decay of the market crash on the other hand."

This paper provides a very interesting approach based on Weierstrass functions (i.e. functions that are continuous everywhere, but differentiable nowhere), managing to recover the Weierstrass function from the renormalisation group equations to construct solutions that are expressed as the combination of a regular and singular part. This extension to the first order is fitted to the observed data in their standard way, providing very accurate fits of this anti-bubble period including the five "crashes" and rebounds that occur as the market drifts lower over time. However, the retroactive predictive ability of these equation do not appear to be impressive, and when one looks at the actual evolution of the market, there is a striking resemblance to the progression of the logistic equation of Lynch & Mestel (2017) and it would be interesting to see how similar extensions of this revised model perform in similar circumstances. The paper concludes with the bold statement that the crashes in this market are all shown to be entirely endogenous including the one following the September 11th terrorist attack on the World Trade Centre in New York.

Continuing with their focus on the 2000 S&P500 anti-bubble, Zhou & Sornette (2005) seemed to abandon the interesting line of enquiry of Zhou & Sornette (2003b) and reverted to the more familiar ground of the first and second order

LPPL models to develop tests to detect log-periodicity regime-switching and locating the end of an anti-bubble. Finally Zhou & Sornette (2006) add a variety of other market factors, such as volatility, interest and currency exchange rates, and conclude that, with all the fundamental indicators available to explain the markets during this period, the model that remains the best descriptor is the second order LPPL model with no additional market factors taken into consideration.

The theoretical development is put to one side at this point, to allow a brief summary the criticism of the claims made by practitioners, particularly Sornette and co-workers.

2.4 Criticisms

The predictions of Gluzman and Yukalov, in particular those contained in their second paper (Gluzman & Yukalov 1998a), where it is explicitly stated that their methods are able to predict market crashes, came under close scrutiny by Laloux et al. (1998). The authors systematically tested the theory presented by Gluzman & Yukalov (1998a) and found the method performed much worse than a compatible null hypothesis in predicting price movements over the period between 1990 and 1998. Johansen & Sornette (1999c) also argued against the logarithmic divergence of Vandewalle, Boveroux, Minguet & Ausloos (1998) in that departing from the generalisation of the power law is not a sensible approach to the critical exponent question nor does it perform as well in analysing market observations; moreover, at the critical point this logarithmic divergent model becomes unbounded and this cannot happen in practice. They also discuss the relative merits of the power law model and the logarithmic model of Vandewalle et al. in practical application and conclude that the logarithmic model does not perform well in such qualitative and quantitative comparison.

Laloux et al. (1998) go on to criticise the *ex ante* predictions of Vandewalle et al. in respect of the October 1997 event. The thrust of the objections raised by the authors (including direct refutation of the results given by Gluzman & Yukalov 1998a), is that all the examples of log-periodicity preceding so called

crashes are no more than "anecdotal", and do not come as support for any strong theoretical argument, nor do they provide any compelling experimental evidence. The authors cast doubt on the evidence of predictability of these methods and accuse proponents of the theory as "lacking scientific rigour".

Some years later, in 2002, there was an exchange of letters between Johansen and Laloux et al. (Johansen 2002, Laloux et al. 2002) in which Johansen responded to the criticisms of Laloux et al. in respect of using the LPPL models as a predictor of impending market crashes. Johansen pointed out that the criticisms were based on an arbitrary rather than systematic methodology to identify phenomenological features not detected by the model and as such their analysis presented a distorted view of the experiments. Laloux et al. replied to Johansen's comments, and conceded that, since the original paper was published, more evidence had been accumulated on the universality of parameters governing crashes, but that the underlying theory was still open to strong criticism as was presented by Feigenbaum (2001b) and Feigenbaum (2001a), which, the authors point out, is rarely cited by Sornette et al. Despite their concession, Laloux et al. continued to express the lack of reporting of the failures of so called "predication", and stressed that the assumption of a crash being outliers to a stretched exponential has no basis in any theory, and cite research into studies that have shown the existence of much fatter tails in the distribution of draw downs where the disputed predicted events would not be seen as outliers.

The underlying assumption of "rational expectation" in the JLS model presented by Johansen & Sornette (1999a) was criticised strongly by Ilinski (1999). Firstly, the author questioned the universality of the empirical evidence of log-periodic precursors to market crashes, as these signals could not be detected prior to any event in the Russian market or other unspecified emerging markets, and pointed out that if the "sell" signals were followed in the more successfully LPPL-fitted market crashes, such a strategy would have led to large losses. Ilinski suggests this is why Johansen & Sornette (1999a) introduced the probabilistic hazard rate model, in his opinion reducing the practical usefulness of the model as a predictive tool. He also points out that the features of log-periodicity has

been known by technical traders for some time and that so called "log-spirals" are used to identify both large and small movements very much in the spirit of Drozdz et al. (1999). Despite also questioning the interconnectedness of financial traders, Ilinski's main objection is that the rational expectation derived from Johansen's assumption of a martingale condition can only mean that the best prediction of the market return should be considered zero, and, in order for a bubble to form, there must be added a term which gives traders the incentive to stay in the market. As this premium is a function of risk, it would also need to be modelled in some way. Johansen et al. (1999) answered the market rationality objection by pointing out that Ilinski's misunderstanding is that the risk premium he speaks about is embedded in the acceleration of returns implied by an increasing hazard rate. Added to this, the authors analysed four crashes in different Russian stock market indices and found evidence of log-periodic precursors all with similar critical exponents and scaling factors in direct contradiction to Ilinski's findings.

The aforementioned, and rarely cited paper by Feigenbaum (2001b) concludes that the log-periodicity of the S&P500 index prior to the 1987 crash is not statistically significant if the time series is truncated one year prior to the actual event, and throws additional doubt on the evidence of such crashes being outliers. Feigenbaum complained that up until then there had been very little in the way of rigorous statistical analysis of the log-periodic precursors to financial crashes, and proponents were focused on the qualitative evidence derived from curve fitting or through spectral analysis. Despite a limited amount of tentative work (Feigenbaum & Freund 1998, Johansen et al. 2000, Johansen et al. 1999) having been done in this vein, Feigenbaum stated that "the null hypothesis that log-periodic spell are the casual products of random fluctuations has *not* been rejected". The author tested log-periodicity in the first difference of the time-series, which according to the JLS model should also be apparent, since this will remove the inherent autocorrelation. His point here was that if one cannot find log-periodicity in the first difference, one could perhaps prescribe the log-periodicity in the price to serial-correlation effects rather than anything else. i.e. the short

term trends driven by autocorrelation give the appearance of log-periodic behaviour purely by chance. His results suggested that the model fits of the entire data between 1980 and 1987 were in fact statistically significant. However if one truncates the data at June 1986 the statistical significance disappears, and there is no apparent log-periodicity in the first difference of the time series. Feigenbaum's argument was that for any value to be ascribed to the JLS model it should have some predictive ability, and if the null hypothesis cannot be rejected for the truncated time-series, then one cannot place value on this model. Feigenbaum made three other points:

1. that the JLS model is founded on a dubious premise and that he agrees somewhat with Ilinski (1999) in that, since the log-periodicity of the price evolution is derived from the herding of highly interconnected irrational traders, and that a single risk-adverse rational agent is included to maintain the no-arbitrage environment (the martingale condition that the expected price in the next instant, given all information is available to all traders prior to this time, is the current price, i.e. the rational expectation is that gains cannot be made) cannot hold without the addition of a risk premium. The point can be conceded if the risk premium is constant, but this is not the case and thus the log-periodicity would be "disrupted" unless risk adverse agents are discounted from the model;

2. in fitting these high-dimensional models, the are many local minima with very similar r.m.s. values so that the choice of parameter value may not be informative. Additionally, given the certainty of serial correlation in the price, the question of standard error is further complicated. Furthermore, there is nothing unusual about finding periods of log-periodicity, as one can find these periods even where no crash occurred. Research based on modelling stock markets with either random walks or GARCH processes (general autoregressive conditional heteroskedasticity models are often used to model time-series with time-varying volatility) that show no evidence of log-periodicity can be discounted as there is no reason that a crash should occur in the models at all; and,

3. the frequency of draw-downs of particular sizes seems to follow an exponential distribution with the major crashes of the century lying off this curve, implying to Sornette et al. that apart from these large events, markets are driven by a GARCH process, and the outliers are governed by herding behaviour. However, there is no reason to take an exponential distribution, or even a stretched exponential. For instance, Feigenbaum shows that a distribution specified by $N(d) = \exp(A + B\sqrt{d} + Cd)$ fits all data points including the large drawdowns with $R^2 = 0.9519$. Feigenbaum concludes from this that there is no evidence of two populations of separate regimes.

Sornette & Johansen (2001) took great issue with Feigenbaum (2001b) and provided a comprehensive historical retrospective and, at some times didactically condescending, analysis of Feigenbaum's findings. Feigenbaum (2001a) responded in strong terms to these criticisms, and the discussion provides a good summary of the main areas of contention along with the arguments and counterarguments. This summary is given here:

Rational expectation paradigm

Sornette et al. criticism: The rational expectation bubble theory is robust because the crash hazard rate $h(t)$ governs the objective risk of the crash occurring, and that the modification of the pricing kernel suggested by Feigenbaum may distort the log-periodic signal but it remains intact, and general form of risk aversion does not invalidate the theory. Moreover, returns conditional on a crash not having occured are non-zero, and as such, risk-averse agents can expect profits.

Feigenbaum response: Firstly, the rational agent cannot base decisions on conditionality and since there cannot be certainty of a crash not happening the rational agent will stop the bubble from forming in the first place. Secondly, Sornette et al.'s, suggestion of inserting a "stochastic discount factor" as the pricing kernel may be valid in allowing all rational agents to be counted as one, but only in markets in equilibrium.

Curve fitting

Sornette et al. criticism: Feigenbaum does not filter out values of the log-frequency, ω, and the critical exponent, β, that should be rejected on the basis of the physical properties of these parameters despite having good r.m.s. or R^2 results, since the errors are not Gaussian in nature, and such statistics are not necessarily the most important factor. This is why the Lomb periodogram has been extensively used, which shows that the parameters β and ω do carry information.

Feigenbaum response: The guidelines on parameter choice and fitting criteria have been chosen to produce the desired result, and this is how the apparent universality in the parameters across a wide range of market crashes has probably come about. Furthermore, the Lomb periodogram is not independent of the curve fit as the detrending function is dependent on some of the parameters.

Statistical significance

Sornette et al. criticism: The fact that the GARCH process produces log-periodicity that *does not* precede a crash further strengthens the idea that a different mechanism is at work during bubble formation, and this is equally applicable to the frequency of crashes. Additionally, Feigenbaum's own results show statistical significance of the log-periodicity at the 95% level when the first difference is analysed. After removing the most relevant piece of data it is not surprising that the statistical significance is not found.

Feigenbaum response: All that can be concluded is that the stochastic process governing stock markets is not a GARCH process. Once cannot conclude that the markets are governed by the JLS model, whereas one cannot reject the null hypothesis that random-walks can produce log-periodic behaviour, and thus such behaviour in the first difference is not extraordinary. Furthermore, the argument that the last year of data preceding the 1987 crash is the most relevant presupposes that critical behaviour is being examined, and this had not actually been established.

Crashes as outliers

Sornette et al. criticism: Reshuffling the daily returns of the NASDAQ index (i.e. maintaining the distribution but removing the correlations) yielded results that the largest four draw-downs had a 0.01% chance of occurring by chance. also, As mentioned above, GARCH models of the NASDAQ were unable to provide any evidence that draw-downs of this size and frequency should arise at random.
Feigenbaum response: One would expect the reshuffling method to produce these results at daily returns are not uncorrelated, and as stated above the GARCH process failing to produce such draw-downs is only evidence that the markets do not follow a GARCH process. Finally, one needs to distinguish between draw downs that do have log-periodic precursors and those that do not, and these differing in a statistically significant manor may point to separate mechanisms being at work.

Feigenbaum (2001a) conclusions is a strong general criticism of Sornette et al. which is worth adding here verbatim:

> "...it is important not to get carried away by a new hypothesis and lose one's scientific objectivity. [Sornette and Johansen] lament that "it is an all too common behavior to dismiss lightly a serious hypothesis by not taking the trouble to learn the relevant skills necessary to test it rigorously." However, they themselves have thrown roadblocks in the way of testing it rigorously. As long as they keep the details of their methodology secret, their results cannot be independently tested or reproduced. They have also disregarded many econometric procedures that would generally be considered the most appropriate tools for addressing the problems they encounter. Instead, they cobble together methods with little or no theoretical basis and then they question why nobody follows their trail. Furthermore, when one finds that standard methods produce results contrary to their hypothesis, one has to consider if maybe they have rejected these standard methods precisely because the results obtained do not support their hypothesis."

Despite the intensity of this exchange, perhaps the most serious challenge to the JLS model came from Chang & Feigenbaum (2006), who claim to be the first group to use Bayesian methods to analyse the statistical significance of the LPPL and JLS models. The conclusions reached by the authors probably made for some uncomfortable reading by the main proponents of the theory, Sornette and his co-workers. The analysis computes the marginal log likelihood, i.e. a measure of how likely the model is true given the observed data, of several different models, namely those where daily returns vary log-periodically versus the null hypothesis of a constant drift for both calendar and market time (i.e. non-trading days are ignored), and those where the probability of the crash occurring also varies log-periodically (as in the JLS model) against the same null hypothesis. The authors came to the conclusion that, at least when one is examining the October 1987 bubble and crash, there is no evidence that the JLS model can explain the log-periodic oscillations.

These conclusions are echoed by Bree & Joseph (2013) who also cast doubt on any of the predictive claims expressed by proponents of the theory. As an added criticism, Bree et al. (2013) show that the sloppiness of the LPPL models (this is discussed Section 4.1 of this report) led to unreliable probabilistic windows for the critical time.

Finally in this section the last word is given to Sornette et al. (2013). The authors welcome healthy criticism in the spirit of debate and furtherance of the research, but say "several serious misconceptions seem to be present within this literature both on theoretical and empirical aspects"

2.5 Recent developments

Research in this field has continued in recent years, and a very good summary of the historical background, fitting methodologies and diagnostic testing of log-periodic power laws can be found in Geraskin & Fantazzini (2013). However, in the context of this research it is important to mention the work contained in Cheah & Fry (2015) as this gives a variation of the stochastic equation (2.6)

which is used as the basis for the concluding research work of this thesis.

The focus of this paper was to determine whether one could detect speculative bubbles in the price of the digital currency, Bitcoin. Since its introduction in 2008, the value of Bitcoin has gone through periods of boom and bust despite how its creators had envisaged in "in-built" stability mechanism by virtue of the process by which new coins were created or "mined". Although these process are not within the scope of this thesis, it is perhaps interesting to consider assets that do not have a fundamental value, which indeed Cheah & Fry (2015) goes on to demonstrate. One could imagine that assets of this type are by their very nature prone to be driven by herding behaviour alone, and as such are not contaminated by exogenous events in the same way that traditional, "real-world" assets are. The authors cite publications that would agree that the bitcoin price is driven by nothing more than market sentiment (for example Dwyer 2015 and Weber 2014), however, since the treatment of Bitcoin from a tax, regulatory and legal perspective is in a state of flux on a jurisdiction by jurisdiction basis, it is difficult to say categorically that prices are solely being driven by the herding behaviour of market participants. The authors conclude that up until the Bitcoin crash of December 2013, there is significant evidence of the market being in a bubble regime, and this is confirmed by cross checking their results with the model suggested by Johansen et al. (2000).

This is an interesting result in itself, but the aspect of the analysis contained in Cheah & Fry (2015) which has influenced the concluding research of this thesis is the alternative formulation of the stochastic equation which is used to describe speculative bubbles. Although, on the face of it these two models seem very similar they present a divergence in how the volatility of the asset returns is treated. This divergence in treatment has allowed a middle ground to be assumed, which has led to a novel methodology to detect the formation of bubble regimes as described in Lynch & Mestel (2019).

It is important to note that in Johansen et al. (2000) the stochasticity enters the model though the jump process alone, and there is a straightforward assumption that the volatility, σ_t (using the notation taken for the time-dependant

volatility throughout this thesis) is taken as zero. Conversely, Cheah & Fry (2015), assumes log-normal returns in the asset price, S_t, such that

$$dS_t = \left(\mu_t + \frac{\sigma_t^2}{2}\right) S_t \, dt + \sigma_t S_t \, dW_t \,, \tag{2.12}$$

where $\mu_t + \sigma_t^2/2$ is the drift coefficient which drives price returns, σ_t is the volatility, and W_t is a standard Wiener process with $E[W_t] = 0$ and $E[W_t^2] = t$. Now assuming a jump process, j_t as in (2.6), the asset price is modified by

$$\tilde{S}_t = S_t(1-\kappa)^{j_t} \,, \tag{2.13}$$

where, as before, κ is the amount by which the markets expects the asset price to fall given as crash event. As can be seen, the probabilistic nature of the model not only occurs in the jump process, but also in the log-normal returns governed by the volatility, σ_t.

As described in Chapter 6, the implication of this formulation is that as a crash becomes more likely, returns excepted by market participant will increase, and somewhat unintuitively, the volatility will be expected to fall. Furthermore, since, as will be seen in §6.2.1, the time dependant volatility is given by

$$\sigma_t^2 = \sigma^2 - (\log(1-\kappa))^2 h_t \tag{2.14}$$

where $\kappa \in [0, 1]$, there is the added problematic consequence that if the hazard function, h_t is assumed to take the form of a power-law, then, as the critical time is approached, $h_t \to \infty$, and $\sigma_t^2 < 0$. Clearly, it is not possible to have negative volatility, and, as such, one must assume that theoretic conditions governing the model break down as the critical time is approached.

Chapter 3

Theoretical foundations

Ideas taken from the fields of statistical physics and dynamical systems have underpinned the development of log-periodic power law models of asset price bubbles in financial markets, and before attempting to derive the first and higher order equations which describe these models, it is important to understand why the particular renormalisation group equations upon which these models are based have been chosen as suitable descriptors of the laws governing financial markets in bubble regimes.

The central tenet of the literature concerned with log-periodic oscillations as a precursor to criticality in financial markets, has been that the oscillatory phenomenon arises due to the underlying mechanics of a market exhibiting some sort of self-similarity over a discrete set of scales. This is the discrete scale invariance that has been described previously in Chapter 2. The questions, therefore, that must be answered are what is meant by self-similarity and discrete scale invariance in a formal sense, and how do log-periodic oscillations arise from these conditions?

In this chapter, these questions are answered in two ways. Firstly, in general sense, by investigating the properties of self-similar fractal geometry (see Mandelbrot (1983) and Barnsley (1988) for general information on fractal geometry), and then secondly, a more specific model of interactions between agents in financial markets is proposed, and by using the analogy from the field of statistical mechanics of interacting spins on a lattice, the renormalisation group equations

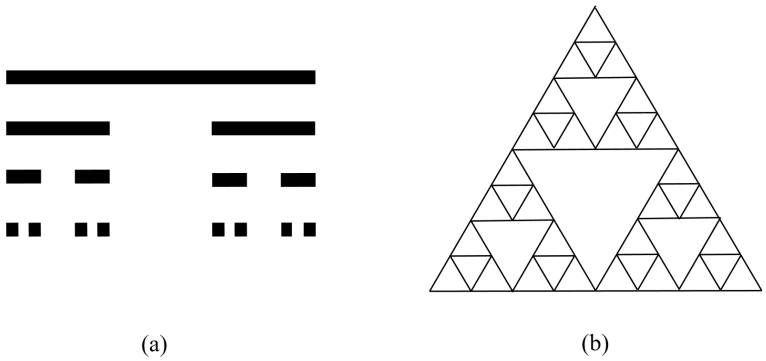


Figure 3.1: (a) Cantor set and (b) the Sierpiński triangle

are recovered. Finally, this leads to the derivation of the first and second order LPPL models.

3.1 Fractal dimension and self-similarity

In very simple terms, a fractal is a set with a non-integer Hausdorff dimension. The formal definition of the Hausdorff dimension is outside the scope of this study, but an interpretation in terms of topological dimension is useful. If, for instance, a line of unit length is taken and partitioned into M equal sections, there will be M similar sections each of size $1/M$. Similarly, if a square of unit area is taken and each side partitioned into M equal sections, this will result in M^2 squares each of area $1/M^2$. Extending this idea to a cube of unit volume, the result will be M^3 cubes of size $1/M^3$. If we denote the number of resulting similar objects from M partitions of side length as N, it can be seen that $N = M^d$, giving the informal definition of the dimension $d = \log M / \log N$. In this context, the Hausdorff dimension may be considered as a measure of the density of self-similarity for a given object.

In this chapter, two fractals that exhibit DSI are taken as examples, the Cantor set and the Sierpiński triangle as shown in Fig. 3.1. The Cantor set is constructed by considering a line segment and dividing it into three sections of equal length. The open middle section is removed leaving two identical lines separated by a gap of the same length. This process of removing the middle third section of each line is continued ad infinitum, thus deriving the Cantor set,

a non-countable and nowhere dense set of points in 1-dimensional space. From the definition above, the line is partitioned in three segments, and the resulting number of line sections is two, therefore, $M = 3$ and $N = 2$ giving a Hausdorff dimension of $d = \frac{\log 2}{\log 3}$.

Similarly, the Sierpiński triangle is constructed by taking an upward pointing equilateral triangle which has a downward pointing equilateral triangle of exactly half the dimensions removed from the centre, thus leaving three similar upward pointing equilateral triangles. Repeating this operation with each of these smaller triangles gives the Sierpiński triangle shown in Fig. 3.1. Clearly, this has a Hausdorff dimension of $d = \frac{\log 3}{\log 2}$.

Taking a more formal approach, generally, fractals may be constructed by considering iterated function systems (**IFS**) (Falconer 2014 and Hutchinson 1981), and specifically, fractals that are self-similar can be constructed as the fixed-point set, F, of an IFS consisting of similarities. A mapping $\phi : \mathbb{R}^n \to \mathbb{R}^n$ is said to be similar if there is an $s > 0$ such that $||\phi(\mathbf{x}_1) - \phi(\mathbf{x}_2)||_2 = s||\mathbf{x}_1 - \mathbf{x}_2||_2$, and an IFS is a finite collection of mappings $\{\phi_1, \phi_2, \ldots, \phi_m\}$, where $\phi_k : \mathbb{R}^n \to \mathbb{R}^n$ and each ϕ_k is a contraction. In what follows, IFSs constructed from mappings which are also similarities are considered, i.e. $0 < s_k < 1$. For such systems the following is known:

1. There exists a unique compact fixed-point set, $\hat{F}$, satisfying $\bigcup_{k=1}^{m} \phi_k(\hat{F}) = \hat{F}$

2. If there exists an open set U with $\phi_k(U) \subset U$ for $k = 1, 2, \ldots m$ and $\phi_i(U) \cap \phi_j(U) = \varnothing$ for $i \neq j$ (Moran's open set condition) then the Hausdorff dimension, $d \geq 0$, is given by the unique solution of the equation

$$\sum_{k=1}^{m} s_k^d = 1 \tag{3.1}$$

The Cantor set can be constructed by the IFS consisting of mappings on $\mathbb{R}$, $\phi_1(x) = \frac{1}{3}x$, and $\phi_2(x) = \frac{1}{3}x + \frac{2}{3}$, where ϕ_1 generates the left most line segment, and ϕ_2 generates the segment on the right. This gives $m = 2$, and since $s = \frac{||\phi_i(a) - \phi_i(b)||_2}{||a-b||_2}$ for $a, b \in \mathbb{R}$ and $i \in (1, 2)$, $s = \frac{1}{3}$. Since Moran's open set

condition is satisfied with $U = (0, 1)$, the Hausdorff dimension for the Cantor set is given by the solution to the equation $\sum_{k=1}^{2} \frac{1}{3^d} = 1$, it can be confirmed that $d = \frac{\log 2}{\log 3}$.

Similarly for the Sierpiński triangle, the IFS on $\mathbb{R}^2$ consists of the mappings $\phi_1(\mathbf{x}) = \frac{1}{2}\mathbf{x}$, which contracts the equilateral triangle of unit base length by a factor of 2, $\phi_2(\mathbf{x}) = \frac{1}{2}\mathbf{x} + \left(\frac{1}{2}, 0\right)$, and $\phi_3(\mathbf{x}) = \frac{1}{2}\mathbf{x} + \left(\frac{1}{4}, \frac{\sqrt{3}}{4}\right)$, which provide the required lateral shifts to the contracted triangle. In this case, $m = 3$ and $s = \frac{1}{2}$, confirming the Hausdorff dimension for the Sierpiński triangle as $d = \frac{\log 3}{\log 2}$. It can be seen that Moran's open set condition is satisfied when the set U is the interior of the equilateral triangle, since any point within this triangle maps uniquely to one of the smaller triangles contained within the original.

Having evaluated the Hausdorff dimension, the origin of the associated log-periodicity can be determined, and one can recover the renormalisation group equations. Firstly, one should note that the Hausdorff dimension can also be defined as being equal to the box (or fractal) dimension. This is defined by $\lim_{r \to 0} \frac{\log N(r)}{-\log r}$, where $N(r)$ is the smallest number of n-dimensional balls of radius r needed to cover the set F. The asymptotics asymptotics of $N(r)$ as $r \to 0$ is given by the following theorem (Falconer 1997):

Theorem 3.1.1. *Let F be the fixed-point set of an IFS consisting of similarities as above, then, as $r \to 0$,*

1. *If $\{\log s_1^{-1}, \log s_2^{-1}, \ldots, \log s_m^{-1}\}$ is a non-arithmetic set, then for some $c > 0$, $N(r) \sim cr^{-d}$*

2. *If $\{\log s_1^{-1}, \log s_2^{-1}, \ldots, \log s_m^{-1}\}$ is a τ-arithmetic set, then for some $c > 0$, $N(r) \sim cr^{-d}p(\log r)$ where p is a positive function of period τ.*

where a τ-arithmetic set is one in which there is a real number τ such that a set of points $\{y_1, y_2, \ldots, y_m\}$ may be expressed as $\{\tau k_1, \tau k_2, \ldots, \tau k_m, \}$ where $k_1, k_2, \ldots, k_m \in \mathbb{N}$ and τ is the largest such number. Conversely, a non-arithmetic set is one which is not τ-arithmetic for any $\tau > 0$.

Taking the Cantor set as an example, it can be seen that since $s_1 = s_2 = \frac{1}{3}$, and the set $\{\log 3, \log 3\}$ being clearly $\log 3$-arithmetic, $N(r) \sim cr^{-\frac{\log 2}{\log 3}}p(\log r)$

as $r \to 0$, with the function p being $\log 3$-periodic. Similarly for the Sierpiński triangle, $N(r) \sim cr^{-\frac{\log 3}{\log 2}} p(\log r)$ as $r \to 0$ with the function p being $\log 2$-periodic.

Now, from a geometrical perspective, the reciprocal of the radius, $x = r^{-1}$, can be interpreted as a "magnification factor" such that $N(x)$ is equal to the number of self-similar objects visible at this magnification, i.e. as the magnification is increased, more self-similar objects become visible. Therefore, as $x \to \infty$, the number of visible self-similar objects is given asymptotically as

$$N(x) \sim cx^d p(-\log x) \tag{3.2}$$

Since, p is a τ-periodic function, it can be seen that

$$\begin{aligned} N(xe^\tau) &\sim c(xe^\tau)^d p(-\log xe^\tau) \\ &\sim c(xe^\tau)^d p(-\log x - \tau) \\ &\sim c(xe^\tau)^d p(-\log x) \end{aligned} \tag{3.3}$$

such that since $p(-\log x) \sim \frac{N(x)}{cx^d}$, it follows that $N(x)$ has a renormalisation group relationship

$$N(x) \sim \frac{1}{\mu} N(\lambda x) \tag{3.4}$$

where $\mu = e^{\tau d}$ and $\lambda = e^\tau$, and the fractal dimension is given by $d = \frac{\log \mu}{\log \lambda}$.

Turning to the Cantor set, it is known that the fractal dimension $d = \frac{\log 2}{\log 3}$, and that $\tau = \log 3$, $\mu = 2$ and $\lambda = 3$ so that

$$N(x) \sim \frac{1}{2} N(3x) \tag{3.5}$$

Now, $N(x) = 2^k$ for $3^k \leq x < 3^{k+1}$, where $k \geq 0$, i.e. the number of visible self-similar objects is constant in this region until $x = 3^{k+1}$. For example, 2^1 self-similar line segments come into view only when the magnification factor moves from 3^0 to 3^1. Similarly, 2^2 self-similar objects will become visible only when the magnification factor becomes 3^2. Therefore (taking the asymptotic resolving to equality), since $c(3^k)^d p(-\log 3^k) = 2^k$, and recalling that p is $\log 3$-periodic, one

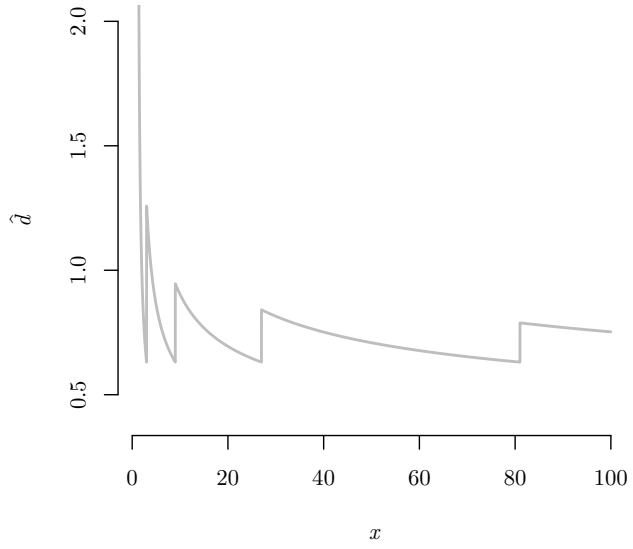


Figure 3.2: The modified fractal dimension of the Cantor set over a continuous range of magnifications equates to the fractal dimension when x is a power of 3.

can see that $c = p(\log 3)^{-1}$ and that

$$N(x)p(\log 3) = x^{\frac{\log 2}{\log 3}} p(-\log x) \tag{3.6}$$

Within the range $3^k \leq x < 3^{k+1}$, $N(x)p(\log 3)$ is a constant, z_k, so allowing $z = -\log x$, in this range

$$z_k = e^{-z\frac{\log 2}{\log 3}} p(z)\,, \tag{3.7}$$

so that the function $p(z)$ can be written

$$p(z) = z_k e^{z\frac{\log 2}{\log 3}}\,. \tag{3.8}$$

However, since it is known that this function, $p(z)$, is $\log 3$ periodic, it can be written more generally as

$$p(z) = z_k e^{\left\{\frac{z}{\log 3}\right\}\log 2} = z_k 2^{\left\{\frac{z}{\log 3}\right\}} \tag{3.9}$$

where $\{\nu\} = \nu - \lfloor\nu\rfloor$. As such, for the Cantor set, the number of self-similar

objects visible at magnification x can be written as

$$N(x) = \frac{1}{2} x^{\frac{\log 2}{\log 3}} 2^{\left\{-\frac{\log x}{\log 3}\right\}} \tag{3.10}$$

Now letting $N(x) = N(1)x^{\hat{d}}$, where $\hat{d}$ is a modified fractal dimension such that $N(x)$ remains constant within the range $3^k \leq x < 3^{k+1}$ for $k = 0, 1, 2, \ldots$, it can be seen that

$$\hat{d} = \frac{\log 2^{\left\{-\frac{\log x}{\log 3}\right\}}}{\log x} + \frac{\log 2}{\log 3} \tag{3.11}$$

Fig. 3.2 shows a plot of the modified fractal dimension and it can be seen that this equates to the fractal dimension when the magnification x is a power of 3, and illustrates how the modified fractal dimension must vary in a log-periodic manner in order for the number of visible self-similar objects to be a whole number.

This analysis for fractal dimension illustrates how discrete scale invariance gives rise to the renormalisation group equation 3.4, and how the modified fractal dimension is log-periodic in nature. The following section shows an alternative way of arriving at the same conclusions vis-à-vis discrete scale invariance analysis of interactions between spins on a diamond lattice as an analogy for interactions between participants in financial markets.

3.2 Two dimensional Potts model

As explained earlier, there have been attempts to explain the apparent log-periodic features presented by financial markets during historical periods of so called bubbles by analogy to models from statistical physics. The model which has become the bedrock for this analogy is the two-dimensional Potts model, which consists of "spins" of microscopic nodes on a crystalline lattice. The interaction of these connected nodes is mapped to the hypothesised imitation between connected agents in financial markets. Although this is a greatly simplified model of such connectivity between agents in financial markets, the two-dimensional Potts model and financial markets do contain some common features. For example, in the Potts model, some nodes are connected to very few neighbouring nodes,

whereas others are connected to a great many. This is a representation how agents in financial markets influence each other. From studying the Potts model, one can recover the renormalisation group equations supposedly governing the log-periodic oscillation in the observed data.

3.2.1 Boltzmann distribution and partition function

When considering models of complex systems it is necessary to understand the probability of such a system being in a particular state within the context of the model. To illustrate the point, imagine a collection of n boxes arranged in a sequential row in which u packets of a fixed amount of energy are distributed. This quantum of energy is denoted as ϵ.

Taking each packet of energy to be a marble, and each wall separating a pair of boxes to be a match-stick, the result of this problem is reduced the finding the number of ways one can arrange the u marbles and $n-1$ match-sticks (Müller 2014). For instance, the are $u+n-1$ locations in which must be placed u marbles. The number of ways this can be done, and therefore, the total number of possible energy distributions is given by

$$\Omega(n,u) = \binom{u+n-1}{u} = \frac{(u+n-1)!}{u!(n-1)!}\,. \tag{3.12}$$

Now, if a single box, A, is considered to be the system in question, and this is joined to a thermal bath[1] represented by n other boxes such that the total system, the combination of A and the thermal bath, is in thermal equilibrium, the question to be answered is what is the probability p_j that A will be in a state such that the energy of A is $E_j = j\epsilon$. If j packets are distributed to A, this reduces the energy packets available to the thermal bath to $u-j$. Therefore the total number of possible energy distributions to the thermal bath is

$$\Omega(j,n,u) = \binom{u-j+n-1}{u-j} = \frac{(u-j+n-1)!}{(u-j)!(n-1)!}\,. \tag{3.13}$$

[1] a system whose temperature remains constant when brought into thermal contact with another system due to its infinite heat capacity.

The probability, p_j, can then be calculated as the ratio of $\Omega(j,n,u)$ to the total number of ways u energy packets can be arranged in the $n+1$ boxes of the total system, i.e.

$$\begin{aligned} p_j &= \frac{\Omega(j,n,u)}{\Omega(n+1,u)} \\ &= \frac{(u-j+n-1)!u!n!}{(u-j)!(n-1)!(u+n)!} \\ &= \frac{n}{u+n}\frac{(u-j+n-1)!u!}{(u-j)!(u+n-1)!} \end{aligned} \tag{3.14}$$

By taking logarithms one arrives at the expression

$$\log\left(p_j\frac{u+n}{n}\right) = \log(u-j+n-1)! + \log u! - \log(u-j)! - \log(u+n-1)! \tag{3.15}$$

and, by using Stirling's formula $\log a! = a\log a - a + O(\log a)$ as $a \to \infty$, the linear terms cancel and the expression becomes

$$\begin{aligned} \log\left(p_j\frac{u+n}{n}\right) \simeq &(u-j+n-1)\log(u-j+n-1) \\ &+ u\log u - (u-j)\log(u-j) \\ &- (u+n-1)\log(u+n-1)\,. \end{aligned} \tag{3.16}$$

Since the thermal bath is not affected by energy fluctuations in the system, the approximation $u-j \approx u$ may be made within the slowly varying logarithms and, as such,

$$\begin{aligned} \log\left(p_j\frac{n}{u+n}\right) \simeq &(u-j+n-1)\log(u+n-1) + u\log u \\ &- (u-j)\log u - (u+n-1)\log(u+n-1) \end{aligned} \tag{3.17}$$

and by further assuming $n \gg 1$ the expression reduces to

$$\begin{aligned} \log\left(p_j\frac{u+n}{n}\right) &\simeq -j\log(u+n) + j\log(u) \\ &= -j\log\left(1+\frac{n}{u}\right)\,. \end{aligned} \tag{3.18}$$

Now, turning to the definition of entropy $S = k_b \log \Omega(n+1, u)$, where k_b is the Boltzmann constant, one may write

$$S = k_b \log \frac{(u+n)!}{u!n!} \tag{3.19}$$

which, on using Stirling's formula, simplifies to

$$\begin{aligned} S &\simeq k_b \left((u+n)\log(u+n) - u\log u - n\log n \right) \\ &= k_b \log \left(\frac{(u+n)^n (u+n)^u}{u^u n^n} \right) \\ &= k_b \log \left(\left(1 + \frac{u}{n}\right)^n \left(1 + \frac{n}{u}\right)^u \right) \\ &= k_b \left(n \log \left(1 + \frac{u}{n}\right) + u \log \left(1 + \frac{n}{u}\right) \right) . \end{aligned} \tag{3.20}$$

Since the reciprocal of the temperature, T, is defined as the partial derivative of the entropy, S, with respect to the internal energy, it can be seen that by treating u as a continuous variable, the internal energy, μ, becomes $\mu = u\epsilon$ and

$$\frac{1}{T} = \frac{\partial S}{\partial \mu} = \frac{\partial S}{\partial u}\frac{du}{d\mu} . \tag{3.21}$$

Furthermore, the partial derivative of the entropy with respect to u is given by

$$\frac{\partial S}{\partial \mu} = \frac{k_b}{\epsilon} \log \left(1 + \frac{n}{u}\right) \tag{3.22}$$

leading to the expression

$$\frac{\epsilon}{k_b T} = \log \left(1 + \frac{n}{u}\right) . \tag{3.23}$$

Therefore, from equation (3.18) with $E_j = j\epsilon$ one obtains the relationship for the probability distribution of the system A being in an energy state E_j.

$$p_j = \frac{n}{u+n} e^{-\frac{E_j}{k_b T}} . \tag{3.24}$$

The probability distribution $e^{-\frac{E_j}{k_b T}}$ is known at the Boltzmann distribution, which is a standard result that can be found in many textbooks on statistical

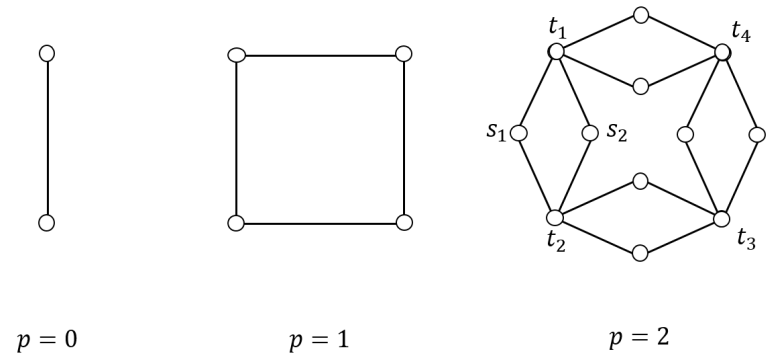


Figure 3.3: The diamond lattice at three degrees of scaling

mechanics. The result (3.24) can be derived much more generally than from starting with (3.12) as has been done here. Since it is a probabilistic certainty that the total probability of finding the system in any one of the energy states is 1, the constant $\frac{n}{u+n}$ must be such that it normalises the Boltzmann distribution for the system, i.e.

$$p_j = \frac{1}{Z} e^{-\frac{E_j}{k_b T}} . \tag{3.25}$$

This normalisation constant, Z is called the partition function and represents all the possible states of such a system when it is in thermal equilibrium with a heat bath at a fixed temperature and is defined as

$$Z = \sum_j e^{-\frac{E_j}{k_b T}} . \tag{3.26}$$

It is by finding an expression for the partition function at different scalings of a lattice model, given a particular coupling factor of adjacent spins, that one may recover the renormalisation group equations which derive the LPPL models. This is the focus of the following sub-section.

3.2.2 Recovering the renormalisation group equations from the diamond lattice

The diamond lattice shown in Fig. 3.3 is invariant under a discrete range of scales as is the Cantor set and Sierpiński triangle. However, in studying the behaviour of financial markets, the diamond lattice model is preferred as it goes

some way to replicating the interactions between market traders; some will have many connections while others may have few.

Firstly, following Saleur et al. (1996), one may make an analysis of how this lattice scales. Referring to Fig. 3.3, with the "scaling index", p, set at 1, there are 4 "sites", connected by 4 bonds. Moving the scaling index up one, such that $p = 2$, it can be seen that the single bond between sites t_1 and t_2 is represented by four bonds connecting the sites via s_1 and s_2. Therefore, at each scaling index, it can be seen that there are $B_p = 4^p$ bonds. The number of sites at scaling index p is $S_p = 2B_{p-1} + S_{p-1}$. Since the number of sites at $S_0 = 2$, this recurrence relationship reduces to

$$\begin{aligned} S_p &= 2\left(\sum_{i=0}^{p-1} B_i + 1\right) \\ &= 2\left(\sum_{i=0}^{p-1} 4^i + 1\right) \\ &= \frac{2}{3}\left(4^p + 2\right) . \end{aligned} \tag{3.27}$$

Assuming that at a particular scaling, p, each vertex has some definable quality, σ_i, referred to as a "spin", where $i = 1, 2, \ldots, Q$, and between neighbouring vertices there is an interaction energy J_p dependent on these spins, then the energy of the system is given by

$$E = J_p \sum_{<i,j>} \delta_{(\sigma_i,\sigma_j)} \tag{3.28}$$

where the sum is taken over all the neighbouring spins on the lattice and $\delta_{(a,b)}$ is the Kronecker delta

$$\delta_{(a,b)} = \begin{cases} 1 & a = b \\ 0 & a \neq b \end{cases} \tag{3.29}$$

As seen in the previous section, the partition function represents all the possible states of the system when it is in thermal equilibrium with a heat bath at a fixed temperature, so to simplify matters, it is assumed that the system under investigation is in such thermal equilibrium. Therefore, the partition function at

a particular scaling for the system is defined as

$$Z_p = \sum_{\sigma} e^{-\frac{E}{k_B T}} \tag{3.30}$$

where the sum is taken over all possible spin configurations.

It is the partition function that, once determined, enables all the thermodynamic quantities of the system to be discovered. Given the number of spin configurations possible in a lattice, this could be a difficult task for a large system, but it can be simplified by focussing attention on a single diamond. As has been seen, in Fig. 3.3 two interconnected spins are marked t_1 and t_2, and the two intermediate spins are marked s_1 and s_2. One can calculate this diamond's contribution to the total partition function at the scaling $p = 2$ by calculating all possible outcomes of $\delta_{(t_1,s_1)}$, $\delta_{(t_1,s_2)}$, $\delta_{(t_2,s_1)}$, and $\delta_{(t_2,s_2)}$. This can be taken in two stages, firstly when $t_1 = t_2$ and secondly when $t_1 \neq t_2$.

The sum in equation (3.28) can take values of $0, 1, 2$ or 4, with 3 being logically impossible, and the analysis proceeds by finding how many ways each of these sum can be achieved. The number of configurations does not depend on the state of t_1 or t_2 but only whether they are equal or not. Taking the case when $t_1 = t_2$ we can see immediately that there is only one way a sum of 4 can be achieved, and this is when $s_1 = s_2$. If, on the other hand, $t_1 \neq t_2$, clearly there is no configuration of s_1 and s_2 that can give 4 as result for the sum.

The total number of state combinations available to s_1 and s_2 are summarised in Fig. 3.4(a). Let t_1 and t_2 be in the state σ_i. It can be seen that for the Kronecker delta sum to be equal to 2, s_1 can take the same state σ_i as t_1 and t_2 and s_2 must take any one of the other $Q - 1$ states. Additionally, s_2 may take the state σ_i, and s_1 may take one of the other $Q - 1$ states. Therefore the total number of combinations for the sum to be equal to two in this case is $2(Q - 1)$. The following table gives the results from Fig. 3.4

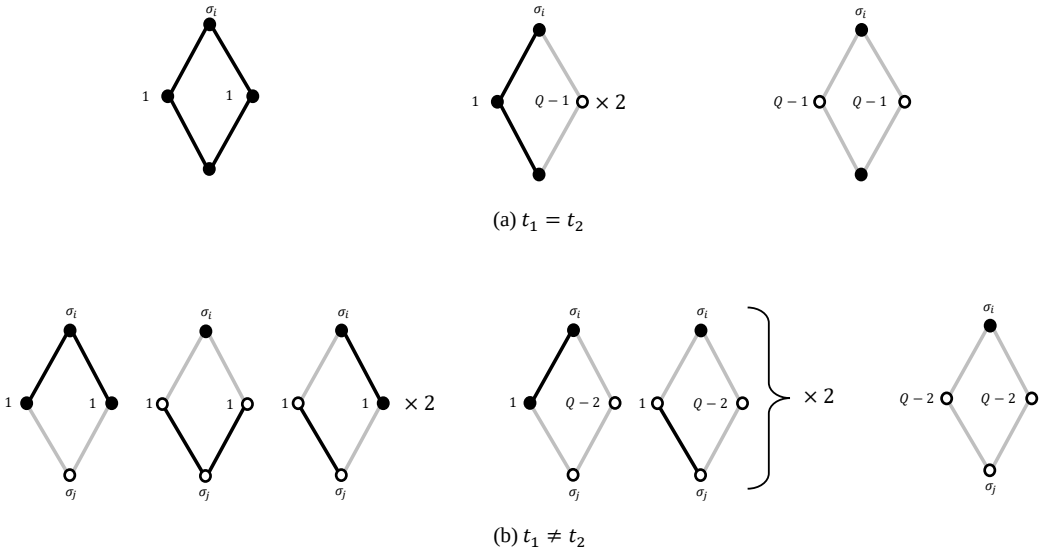


Figure 3.4: Number of state combinations of s_1 and s_2 for Kronecker delta sum to be equal to $4, 2, 1$, and 0

Sum total	$\mathbf{t_1 = t_2}$	$\mathbf{t_1 \neq t_2}$
4	1	0
2	$2(Q-1)$	4
1	0	$4(Q-2)$
0	$(Q-1)^2$	$(Q-2)^2$

Turning now to the partition function for this scaling, Z_p, combining (3.28) and (3.30), the equation for the partition function becomes

$$Z_p = \sum_{\sigma} e^{-\frac{J_p}{k_B T} \sum_{<i,j>} \delta_{(\sigma_i, \sigma_j)}} \tag{3.31}$$

where the sum is taken aver all possible states. If one allows $K_p = e^{-\frac{J_p}{k_B T}}$ then,

$$Z_p = \sum_{\sigma} K_p^{\sum_{<i,j>} \delta_{(\sigma_i, \sigma_j)}} \tag{3.32}$$

and, as such, it can be seen from the table above that the contribution of this

diamond to the partition function can be expressed as

$$Z_p^{(t_1,t_2)} = \begin{cases} K_p^4 + 2(Q-1)K_p^2 + (Q-1)^2 = (K_p^2 + Q - 1)^2 & t_1 = t_2 \\ 4K_p^2 + 4(Q-2)K_p + (Q-2)^2 = (2K_p + Q - 2)^2 & t_1 \neq t_2 \end{cases} \tag{3.33}$$

However, since this piecewise equation depends only on $\delta_{(t_1,t_1)}$, it may be combined into a single expression as follows

$$Z_p^{(t_1,t_2)} = (2K_p + Q - 2)^2 \left(1 + \left(\frac{(K_p^2 + Q - 1)^2}{(2K_p + Q - 2)^2} - 1\right)\delta_{(t_1,t_2)}\right). \tag{3.34}$$

Now, by letting

$$K_p' = \left(\frac{K_p^2 + Q - 1}{2K_p + Q - 2}\right)^2, \tag{3.35}$$

it can be seen that since $\delta_{(t_1,t_2)}$ is either 1 or zero, $\left(K_p'\right)^{\delta_{(t_1,t_2)}} = (1 + (K_p' - 1)\delta_{(t_1,t_2)})$ so that one may write

$$Z_p^{(t_1,t_2)} = (2K + Q - 2)^2 {K_p'}^{\delta_{(t_1,t_2)}}. \tag{3.36}$$

By extension, one may perform this sum over each of the other three diamonds (rightmost in Fig 3.3) and deduce the partition function, Z_p to be

$$\begin{aligned} Z_p &= \sum_\sigma Z_p^{(t_1,t_2)} Z_p^{(t_2,t_3)} Z_p^{(t_3,t_4)} Z_p^{(t_4,t_1)} \\ &= \left((2K_p + Q - 2)^2\right)^{4^{p-1}} \sum_\sigma {K_p'}^{\delta_{(t_1,t_2)} + \delta_{(t_2,t_3)} + \delta_{(t_3,t_4)} + \delta_{(t_4,t_1)}}. \end{aligned} \tag{3.37}$$

One can include the expression for the partial sum $Z_p^{(t_1,t_2)}$ in this manner since the sum is a sum over all spins, and the intermediate spins of s_1 and s_2 are integrated out. From (3.32), the partition function at the scaling $p-1$ can be written as

$$Z_{p-1} = \sum_\sigma K_{p-1}^{\delta_{(t_1,t_2)} + \delta_{(t_2,t_3)} + \delta_{(t_3,t_4)} + \delta_{(t_4,t_1)}}. \tag{3.38}$$

By inspection, one can find that the recursive assignment

$$K_{p-1} \equiv \left(\frac{K_p^2 + Q - 1}{2K_p + Q - 2} \right)^2 \equiv K'_p \equiv \phi(K_p) \tag{3.39}$$

should be allowed so that

$$Z_{p-1} = \sum_{\sigma} \phi(K_p)^{\delta_{(t_1,t_2)}+\delta_{(t_2,t_3)}+\delta_{(t_3,t_4)}+\delta_{(t_4,t_1)}} \tag{3.40}$$

The free energy of the system is the same no matter at which scaling one looks, therefore, the partition function must also be the same when looked at on each scale, and the only quantity which varies is the coupling factor, K_p. As such, from equations (3.37) and (3.40) we have the relationship

$$Z_p(K_p) = Z_p(K_p)((2K_p + Q - 2)^2)^{4^{p-1}} Z_{p-1}(\phi(K_p))\,. \tag{3.41}$$

Now, the free energy, f_p, which is defined as the negative logarithm of the partition function per bond, $f_p(K_p) = -\log Z_p(K_p)/4^p$, which from equation (3.41) gives

$$\begin{aligned} f_p(K_p) &= -\frac{1}{4^p} \log Z_{p-1}(\phi(K_p)) - \frac{4^{p-1}}{4^p} \log(2K_p + Q - 2)^2 \\ &= g(K_p) + \frac{1}{4} f_{p-1}(\phi(K_p)) \end{aligned} \tag{3.42}$$

where $g(K_p) = -\frac{1}{4}\log(2K + Q - 2)^2$

Therefore, for an infinite fractal which has a microscopic coupling of J we can write

$$f(K) = g(K) + \frac{1}{\mu} f(\phi(K)) \tag{3.43}$$

and, as such, the renormalisation group equation that governs the LPPL models for financial markets, given in equation (3.44) and analysed in the following section, is recovered.

3.3 Log-periodic power law model hierarchy

The work of Sornette et al. (1996) and Feigenbaum & Freund (1996), which spawned the ubiquitous LPPL models for bubble formation in financial markets, had their origins in phenomenological observations of the S&P500 index in the period leading up to the crash of October 1987 and a similar time series progression in the Dow Jones index preceding the great crash of 1929.

This section takes a step back from the evolution of these ideas over the past twenty-five years and revisits the initial renormalisation group formalism from which a series of models have been derived to fit the observed data. These models can be thought of as an attempt to describe the periodic fluctuations around a faster than exponential development in prices observed in financial time series prior to many of the significant market collapses. If a constant-return hypothesis is assumed, one expects to see exponential price development, but, during periods in which the market price is in an accelerating regime prior to a large drawdown, price development can be observed to follow a power law. It has long been suggested (see, for example, Sornette et al. 1996) that this unsustainable growth up until some critical time is a consequence of the tendency for traders or market professionals to exhibit imitative or cooperative behaviour. Furthermore, if analogies of studies of hierarchical systems in the natural sciences (Sornette 2002 and Sornette 2009) are to be extended to financial markets, it would be reasonable to expect observations of discrete scale invariance around this critical time (Feigenbaum & Freund 1996), implying log-periodic corrections to the power law, which are indeed observed in some instances. This fundamental expectation has been the genesis of the last twenty-five years of discussion on the matter of financial crashes as critical points (Geraskin & Fantazzini 2013).

The original LPPL model (which is referred to here as the first-order model) was extended in Sornette & Johansen (1997) to take account of an apparent frequency shift in the log-periodic oscillation observed in the S&P 500 index from the beginning of 1980 to the crash of 1987. This was achieved by considering a Landau expansion and the model was further extended in Johansen & Sornette (1999b) by considering the next order of this expansion. This completes what is

referred to in this thesis as the original LPPL hierarchy of models.

In this section an outline of the LPPL model hierarchy is given, starting with the renormalisation group equation itself, and notation is introduced which follows the notation used by Sornette and co-workers in their development of the theory.

3.3.1 Linear scaling

Let $I(t)$ be a stock index at time t. As explained earlier, in practice, $I(t)$ is frequently taken to be the logarithm of a stock index, thereby removing any constant rate of return trend, but the theory given here applies equally to both cases. The critical time is denoted t_c and the time to criticality by x. If an asset bubble (for which the critical time is in the future and $I(t)$ is trending upwards) is being considered, one may write $x = t_c - t$ and define the difference between the asset log-price at the critical time and the asset log-price at time t, as $F(x) = I(t_c) - I(t)$. For an anti-bubble (for which the critical time is in the past and $I(t)$ is trending downwards [2]), $x = t - t_c$ and $F(x) = I(t_c) - I(t)$. This enables the renormalisation group equations to be written in a standard form which is applicable to both bubbles and anti-bubbles. Note that the critical time corresponds to $x = 0$ in both cases, and one may take $x = |t_c - t|$, whist further noting that for bubbles $t_c > t$ and similarly for anti-bubbles $t_c < t$.

The derivation of the basic power-law model of a stock index log-price $I(t)$ at time t is founded on three assumptions:

1. there is some critical time $t = t_c$ at which point $I(t)$ is singular in its derivative and $I(t_c) \neq 0, \infty$;

2. the value of the index at time t is related to the value of the index at some other time, which in the x coordinate system is denoted $\phi(x)$, and;

[2] this view of an anti-bubble has been adopted generally in the literature and is slightly counter intuitive in that the critical time occurs in the past. However, anti-bubbles should not be thought of as the opposite of asset bubbles, i.e. a market which shows accelerating drawdowns, but rather as a market that exhibits an initial rapid drawdown which decelerates over time. It is for this reason that the critical time is deemed to happen before the initial drawdown.

3. at $t = t_c$, $F(x)$ is invariant with some appropriate scaling, under the transformation $\phi(x)$, i.e. $F(x)$ exhibits nonlinear discrete scale invariance.

This leads to the renormalisation group approach to financial bubbles, for which close to the critical time, $x = 0$, the real function $F(x)$ satisfies the renormalisation group equation:

$$F(x) = g(x) + \frac{1}{\mu}F(\phi(x))\,, \quad \phi(x) = \lambda x + O(x^2)\,, \quad x > 0\,, \tag{3.44}$$

where λ and μ are positive constants and there is a non-singular element to $F(x)$ characterised by some continuous and differentiable function $g(x)$. The function $F(x)$ satisfies $F(0) = 0$ with $F'(x)$ singular at $x = 0$, and as such $g(0) = 0$. The solution set of (3.44) depends significantly on the smoothness class of $\phi(x)$ and $F(x)$. In relation to stock-market indices it is usual to assume that $\phi(x)$ is a differentiable function near to $x = 0$, with $\phi(0) = 0$, $\phi'(0) = \lambda > 0$, and that F is continuous at $x = 0$. Defining $\phi^n(x)$ by $\phi\left(\phi^{n-1}(x)\right)$ for $n \geq 2$, (3.44) can be expanded to give

$$F(x) = g(x) + \frac{1}{\mu}\left(g(\phi(x)) + \frac{1}{\mu}F\left(\phi^2(x)\right)\right) \tag{3.45}$$

which, since $\phi^0(x) = x$, resolves to the general solution

$$\begin{aligned} F(x) &= \sum_{i=0}^{\infty} \frac{1}{\mu^i} g\left(\phi^i(x)\right) \\ &= g(x) + \sum_{i=1}^{\infty} \frac{1}{\mu^i} g\left(\phi^i(x)\right) \\ &= g(x) + \frac{1}{\mu}\sum_{i=0}^{\infty} \frac{1}{\mu^i} g\left(\phi^{i+1}(x)\right) \\ &= g(x) + \frac{1}{\mu}F(\phi(x)) \end{aligned} \tag{3.46}$$

and since x is close to zero, $g(x)$ can be ignored and equation (3.44) can be written as

$$F(x) = \frac{1}{\mu}F(\phi(x))\,. \tag{3.47}$$

Now suppose ϕ is linear so that $\phi(x) = \lambda x$, one can readily verify that a solution to the functional equation (3.47) is $F(x) = kx^{\alpha}$ where both k and α are real valued constants. Substituting this expression into the renormalisation group equation gives $\lambda^{\alpha}/\mu = 1$. However, if we allow $\alpha \in \mathbb{C}$ then this relationship becomes $\lambda^{\alpha}/\mu = e^{i2\pi n}$ where $n \in \mathbb{Z}$, giving

$$\alpha = \frac{\log \mu}{\log \lambda} + \frac{i2\pi n}{\log \lambda}\,. \tag{3.48}$$

Now, since equation 3.47 is a linear equation in F, the general solution can be expressed as the sum of all solutions

$$F(x) = \sum_{n=-\infty}^{\infty} k_n x^{\beta + i\frac{2\pi n}{\log \lambda}} \tag{3.49}$$

where $\beta = \log \mu / \log \lambda$ and $k_n \in \mathbb{C}$. Now writing $x^s = e^{s \log x}$, (3.49) can be written as a Fourier series

$$\begin{aligned} F(x) &= x^{\beta} \sum_{n=-\infty}^{\infty} k_n e^{i\frac{2\pi n}{\log \lambda}\log x} \\ &= x^{\beta} \sum_{n=-\infty}^{\infty} k_n \left(\cos\left(\frac{2\pi n}{\log \lambda}\log x\right) + i \sin\left(\frac{2\pi n}{\log \lambda}\log x\right)\right)\,. \end{aligned} \tag{3.50}$$

Since $F(x)$ is a real valued function, by letting $k_n = |k_n|\, e^{i\psi_n}$, it can be seen that the real part of equation (3.50) can be written as

$$\begin{aligned} F(x) &= x^{\beta} \sum_{n=-\infty}^{\infty} |k_n| \left(\cos\left(\frac{2\pi n}{\log \lambda}\log x\right)\cos\psi_n - \sin\left(\frac{2\pi n}{\log \lambda}\log x\right)\sin\psi_n\right) \\ &= x^{\beta} \sum_{n=-\infty}^{\infty} |k_n| \cos\left(\frac{2\pi n}{\log \lambda}\log x + \psi_n\right) \end{aligned} \tag{3.51}$$

and taking the the terms where $n = 0$ and $n = 1$ gives

$$F(x) = |k_0|\, x^{\beta}\cos\psi_0 + |k_1|\, x^{\beta}\cos\left(\frac{2\pi}{\log \lambda}\log x + \psi_1\right) \tag{3.52}$$

so that by aggregating the constants in equation and substituting $x = |t_c - t|$ the

expression for $F(x)$ becomes

$$F(x) = -B\left|t_c - t\right|^{\beta}\left(1 + C\cos\left(\omega\log\left|t_c - t\right| + \phi\right)\right) \tag{3.53}$$

where $\omega = 2\pi/\ln\lambda$ and ϕ aggregates the terms in ψ_0 and ψ_1. Now, since $F(x) = I(t_c) - I(t)$, by taking $A = I(t_c)$ an expression for $I(t)$ can be written as

$$I(t) = A + B\left|t_c - t\right|^{\beta}\left(1 + C\cos\left(\omega\log\left|t_c - t\right| + \phi\right)\right). \tag{3.54}$$

Following Filimonov & Sornette (2013), the nonlinear term ϕ can be dispensed with and the first order model can be expressed as

$$I(t) = A + B\left|t_c - t\right|^{\beta}\left(1 + C_1\cos\left(\omega\log\left|t_c - t\right|\right) + C_2\sin\left(\omega\log\left|t_c - t\right|\right)\right) \tag{3.55}$$

Reducing the number of non-linear terms in this manner has some important implications in fitting these LPPL models to observed data, and this is discussed further in Chapter 4.

3.3.2 Nonlinear corrections to scaling

Sornette & Johansen (1997) introduces a nonlinear correction to scaling in the renormalisation group equation driven by an apparent frequency shift in the first-order model when fitted to the time series of the S&P 500 index between 1980 and the crash of 1987 (Feigenbaum & Freund 1996). This is illustrated in Fig. 3.6 where Feigenbaum & Freund (1996) fits of (3.55) starting from 1980 and from 1986 up until three weeks before the 1987 crash are shown. By fixing t_c to be the date of the crash, they found values of the angular log-frequency ω of 12.94 and 8.06 respectively. It should be noted that these fits have been made using absolute index price rather than the more usual use of log prices. Moreover, one can see that the fit to the time series from the beginning of 1980 seems to over-oscillate in the period between 1986 and the crash in late 1987.

To account for this apparent frequency shift, Sornette & Johansen (1997) developed a nonlinear correction to the first-order model motivated by Landau

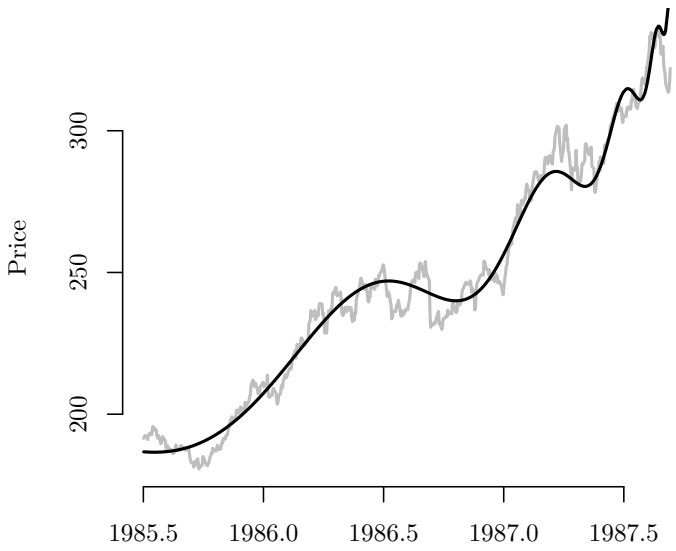


Figure 3.5: Best fit of (3.55) against the S&P500 index price between July 1985, and the crash of October 1987 using nonlinear parameters determined in Sornette et al. (1996). This plot has been created using using $t_c = 1987.74$, $\omega = 7.4$, and $\beta = 0.33$ in conjunction with (4.4). From a visual inspection, one can certainly see that the fit follows the gross features of the price movement during this period. It should be noted that the statistical fluctuations around the best fit has not been analysed.

expansions from the theory of phase transitions in statistical mechanics. Firstly, letting $\nu = \log x$ equation (3.49) can be written as

$$F(x) = \sum_{n=0}^{\infty} k_n e^{\nu\left(\beta + i\frac{2\pi n}{\log\lambda}\right)} . \tag{3.56}$$

Differentiating this expression with respect to ν gives

$$\begin{aligned} \frac{dF(x)}{dv} &= \sum_{n=0}^{\infty} k_n \left(\beta + i\frac{2\pi n}{\log\lambda}\right) e^{\nu\left(\beta + i\frac{2\pi n}{\log\lambda}\right)} \\ &= \sum_{n=0}^{\infty} k_n \left(\beta + i\frac{2\pi n}{\log\lambda}\right) x^{\beta} e^{i\frac{2\pi n}{\log\lambda}\log x} \end{aligned} \tag{3.57}$$

such that

$$\frac{dF(x)}{d\log x} = \alpha F(x) . \tag{3.58}$$

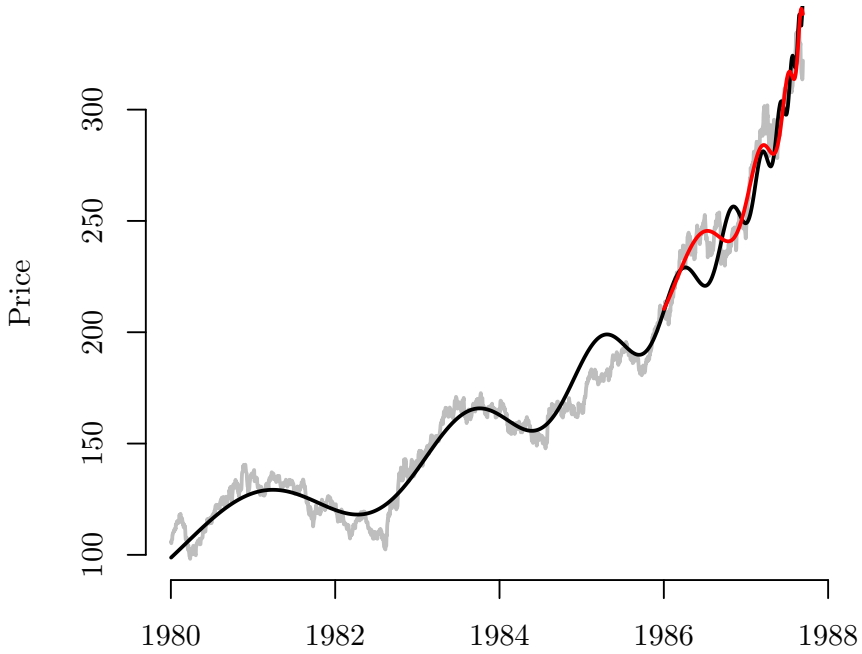


Figure 3.6: Feigenbaum & Freund (1996) first order LPPL fit for S&P500 between 1980 and 1987. The authors assumed that the critical time, t_c, should be equal to the time of the crash. However, this being the case, the angular frequency of the best fitting first order model of the longer term time series (solid black line) implies more rapid oscillations for the period post 1986 than those found by the best fitting model of the sorter time period (solid red line). By holding t_c to be the time of the crash it is not possible to find a fit that satisfies the log periodic oscillations across the whole period. This is shown in Chapter 4, and it is demonstrated that this can be done if one move the value of t_c further out from the actual time of the crash.

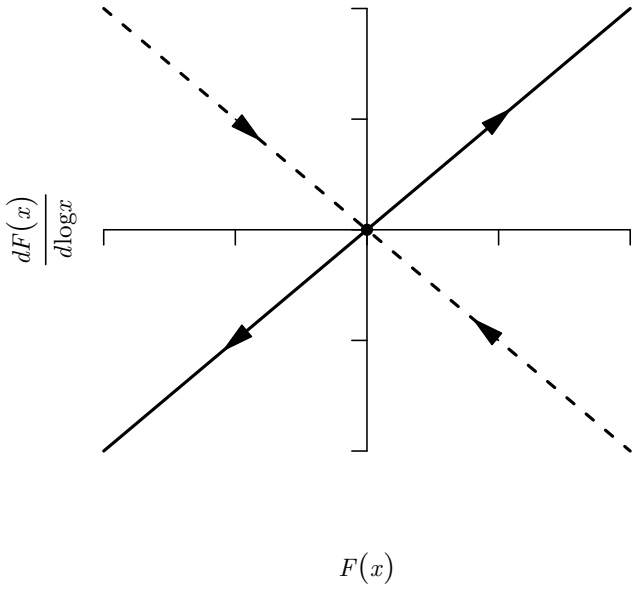


Figure 3.7: Phase plot of the linear differential equation (3.58). When α is positive (solid line), $F(x)$ increases with $\log x$ when $F(x) > 0$, and decreases with $\log x$ when $F(x) < 0$ giving an unstable equilibrium point at $F(x) = 0$. When $\alpha < 0$ (dashed line) the equilibrium point is stable.

Clearly, for real $\alpha > 0$, $|F(x)|$ increases with $\log x$, and conversely when $\alpha < 0$, $|F(x)|$ decreases with $\log x$. Therefore, $F(x) = 0$ is an unstable equilibrium when α is positive, and a stable equilibrium when α is negative. Therefore, regarding α as a bifurcation parameter, the point $\alpha = 0$ is a bifurcation point for a saddle-node bifurcation. Complexifying, writing $\alpha = \beta + i\omega$, with real parameters β and $\omega > 0$, and taking an unfolding of the bifurcation to order $O(|F|^3)$, extending the model such that the bifurcation has a stable region away from this critical point, leads to the following equation as suggested by Sornette & Johansen (1997):

$$\frac{dF(x)}{d\log x} = (\beta + i\omega)\, F(x) + (\eta + i\kappa)\, |F(x)|^2\, F(x) + O(F(x)^5)\,. \tag{3.59}$$

Writing $F(x) = B(x)e^{i\psi(x)}$ such that F is expressed terms of its amplitude B and phase ψ, a relationship for $dF(x)/d\log x$ may be found by again making the

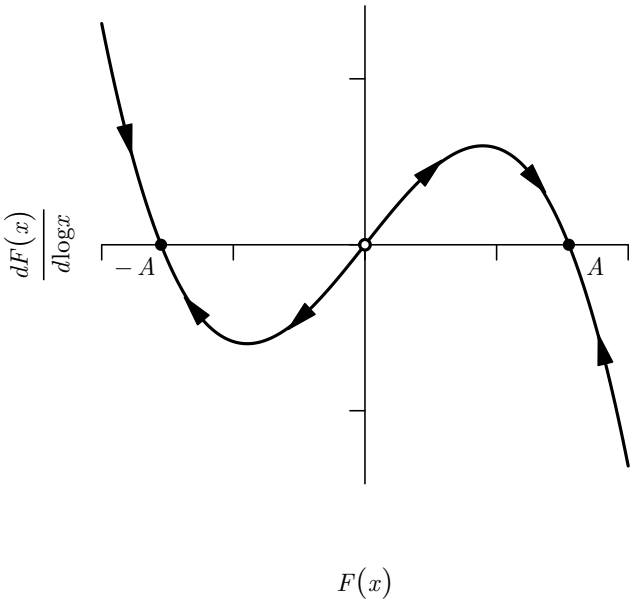


Figure 3.8: With the additional unfolding of the bifurcation to $O(|F|^3)$ the phase plot shows an unstable equilibrium point at $F(x) = 0$, i.e. at the critical time, and two stable equilibria some way away from this point. This seems to fit with the visceral expectation that stable regions of market operation should exist away from the instability of the region around the critical time. This expectation is contemplated by the nonlinear correction.

substitution $\nu = \log x$ to give

$$F(x) = B(e^{\nu})e^{i\psi(e^{\nu})} \tag{3.60}$$

which may be differentiated as

$$\begin{aligned}\frac{dF(x)}{d\nu} &= e^{i\psi(e^{\nu})}\frac{d}{d\nu}(B\,(e^{\nu})) + B(e^{\nu})\frac{d}{d\nu}\left(e^{i\psi(e^{\nu})}\right) \\ &= e^{i\psi(e^{\nu})}\frac{d}{d\nu}(B\,(e^{\nu})) + iB(e^{\nu})e^{i\psi(e^{\nu})}\frac{d}{d\nu}\left(\psi(e^{\nu})\right)\,,\end{aligned} \tag{3.61}$$

which can be written as

$$\frac{dF(x)}{d\log x} = e^{i\psi(x)}\left(\frac{dB(x)}{d\log x} + iB(x)\frac{d\psi(x)}{d\log x}\right)\,. \tag{3.62}$$

Taking equation (3.59) to $O(F(x)^3)$ and substituting the definition of $F(x)$ from

equation (3.60) gives

$$\begin{aligned}\frac{dF(x)}{d\log x} &= (\beta+i\omega)\,B(x)e^{i\psi(x)} + (\eta+i\kappa)\left|B(x)e^{i\psi(x)}\right|^2 B(x)e^{i\psi(x)} \\ &= e^{i\psi(x)}\left((\beta+i\omega)\,B(x) + (\eta+i\kappa)\,B(x)^3\right) \\ &= e^{i\psi(x)}\left(\left(\beta B(x)+\eta B(x)^3\right) + iB(x)\left(\omega+\kappa B(x)^2\right)\right)\,. \end{aligned} \tag{3.63}$$

Equating equations (3.61) and (3.63) results in the differential equations

$$\frac{dB(x)}{d\log x} = \beta B(x) + \eta B(x)^3 \tag{3.64}$$

$$\frac{d\psi(x)}{d\log x} = \omega + \kappa B(x)^2\,. \tag{3.65}$$

The solution of equation (3.64) is

$$x = c_1\left(\frac{B(x)^2}{\beta+\eta B(x)^2}\right)^{\frac{1}{2\beta}} \tag{3.66}$$

where c_1 is a constant of integration, and on rearranging, an expression for $B(x)^2$ is found

$$\begin{aligned} B(x)^2 &= \frac{\beta\left(\frac{x}{c_1}\right)^{2\beta}}{1-\eta\left(\frac{x}{c_1}\right)^{2\beta}} \\ &= \frac{-\frac{\beta}{\eta}\left(\frac{x}{x_0}\right)^{2\beta}}{1+\left(\frac{x}{x_0}\right)^{2\beta}} \qquad \text{where } x_0 = \frac{c_1}{(-\eta)^{-\frac{1}{2\beta}}}\,. \end{aligned} \tag{3.67}$$

The stable equilibrium points are located at $B(x)^2 = -\beta/\eta$ which we presume are at a distance from the unstable equilibrium point at $x = 0$, so by letting $B_\infty^2 = -\beta/\eta$ one can write

$$B(x)^2 = B_\infty^2 \frac{\left(\frac{x}{x_0}\right)^{2\beta}}{1+\left(\frac{x}{x_0}\right)^{2\beta}}\,. \tag{3.68}$$

By substituting this expression for $B(x)^2$ into the differential equation (3.65) it

can be seen that

$$\frac{d\psi(x)}{d\log x} = \omega + \kappa B_\infty^2 \frac{\left(\frac{x}{x_0}\right)^{2\beta}}{1+\left(\frac{x}{x_0}\right)^{2\beta}} \tag{3.69}$$

which has the solution

$$\psi(x) = \omega \log x - \frac{\kappa}{2\beta} B_\infty^2 \log\left(1+\left(\frac{x}{x_0}\right)^{2\beta}\right) + \sigma \tag{3.70}$$

where the constant of integration is σ. Now, by letting $\Delta_\omega = \kappa B_\infty^2$ and $\Delta_t = x_0$

$$B(x) = \frac{B_\infty}{\Delta_t^\beta} \frac{x^\beta}{\sqrt{1+\left(\frac{x}{\Delta_t}\right)^{2\beta}}} \tag{3.71}$$

$$\psi(x) = \omega \log x - \frac{\Delta_\omega}{2\beta} \log\left(1+\left(\frac{x}{\Delta_t}\right)^{2\beta}\right) + \sigma\,. \tag{3.72}$$

These equations for the amplitude and the phase of $F(x)$ are nonlinear corrections or modifications of the linear differential equation (3.58). In this case the amplitude was x^β and the phase was $\omega \log x$. Therefore, we modify this amplitude and phase in equation (3.55) to give the second order model

$$I(t) = A + B\frac{|t_c - t|^\beta}{\sqrt{1+\left(\frac{|t_c-t|}{\Delta_t}\right)^{2\beta}}}(1 + C_1 \cos\ \psi(x) + C_2 \sin\psi(x)) \tag{3.73}$$

where B_∞/Δ_t^β is absorbed into the constant B, and ϕ is absorbed into the constants C_1 and C_2.

One can see that as t approaches the critical time, the angular log-frequency shifts from $\omega + \Delta_\omega \to \omega$, and that the timing of the frequency cross-over is controlled by Δ_t. Another important feature is that as t moves away from the critical point, t_c, the amplitude $B(x)$ saturates to B_∞, i.e. the power law dissipates at large distances from t_c.

By taking higher order unfoldings of the bifurcation, one can obtain a hierarchy of models. For example, an unfolding up to order $O(|F|^5)$ leads to the LPPL third-order model (Johansen & Sornette 1999b). However, as the order

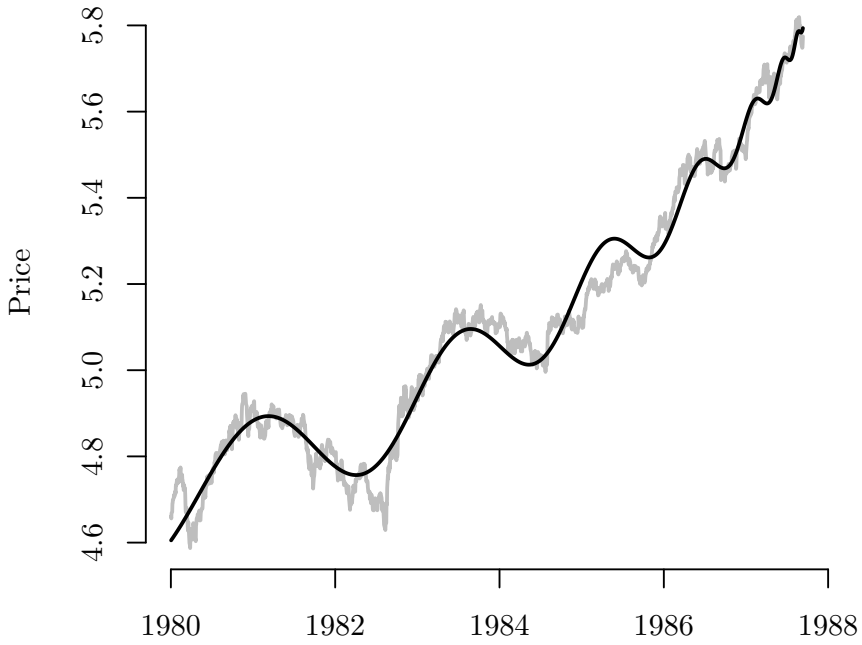


Figure 3.9: Plot of the best fit of (3.73) against the S&P500 Index from 1980 to the crash of October 1987 with nonlinear parameters taken from Sornette & Johansen (1997). In this example, even though the critical time is close to the actual time of the crash, $t_c = 1987.81$, the model fits the gross oscillatory features across the whole time series very well.

Table 3.1: LPPL hierarchy

Model	Amplitude	Phase
Power	x^{β}	
First	x^{β}	$\omega \log x$
Second	$\frac{x^{\beta}}{\left(1+\left(\frac{x}{\Delta_t}\right)^{2\beta}\right)^{1/2}}$	$\omega \log x + \frac{\Delta_\omega}{2\beta} \log\left(1+\left(\frac{x}{\Delta_t}\right)^{2\beta}\right)$
Third	$\frac{x^{\beta}}{\left(1+\left(\frac{x}{\Delta_t}\right)^{2\beta}+\left(\frac{x}{\Delta'_t}\right)^{4\beta}\right)^{1/2}}$	$\omega \log x + \frac{\Delta_\omega}{2\beta} \log\left(1+\left(\frac{x}{\Delta_t}\right)^{2\beta}\right) + \frac{\Delta'_\omega}{4\beta} \log\left(1+\left(\frac{x}{\Delta'_t}\right)^{4\beta}\right)$

increases so does the complexity of solving the associated differential equation and of fitting the resulting model. Table 3.1 shows the four models in the LPPL hierarchy.

Chapter 4

Fitting log-periodic power laws to financial time series

Fitting LPPL models to observed data of financial time series is not a trivial matter, particularly from the perspective of minimising the sum of squares of the residuals. Not only is it a difficult computational task, it is unclear by what measure a set of parameter choices could be described as providing a "best fit". Bree et al. (2013) gives a very interesting account of the challenges faced, in that these models are highly nonlinear and are intrinsically "sloppy". In this context, the term sloppiness is used to describe circumstances where altering some combination of parameters have little effect on the r.m.s. errors whereas small changes in other combination of parameters have very large effects.

It is also a fact that it is very difficult to have confidence that there exists a global minimum with a set of parameters that have any particular meaning, i.e. one can find a set of very similar r.m.s errors which are local minima with wildly different parameter choices, so why should one minimum be chosen over another? Additionally, the choice of data set (i.e. start and end points) can have a significant impact on the minimising parameter choices, particularly on that parameter which is most interesting, the critical time, t_c. As can be seen from Chapter 2, these problems has been discussed extensively in the literature.

One way to reduce the sloppiness of the model (Filimonov & Sornette 2013) is to reduce the number of linear parameters from 4 to 3, i.e. switching from

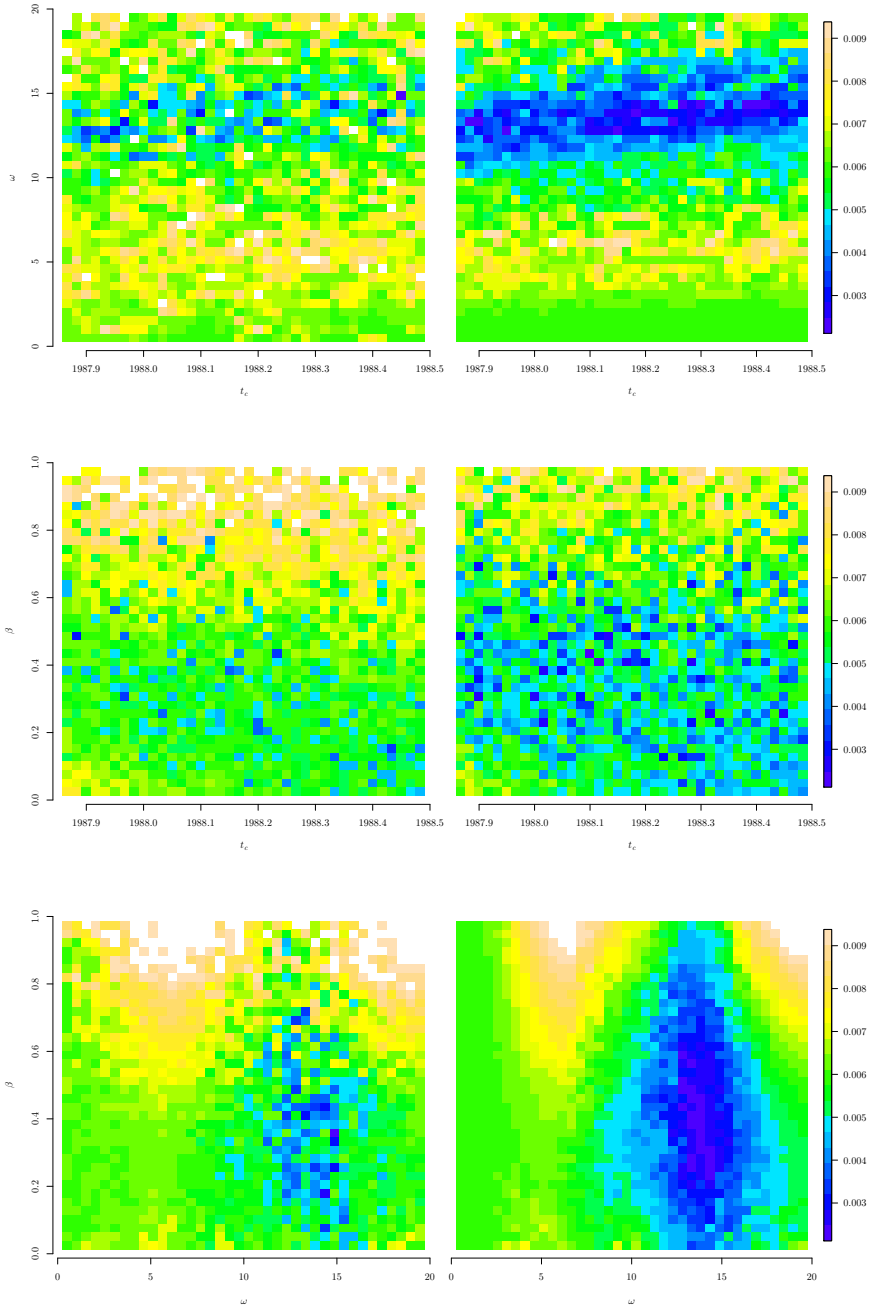


Figure 4.1: The "sloppiness" of the first order LPPL model, with the r.m.s error shown as a heat map for combinations of nonlinear parameter pairs. This example is taken from the 4 nonlinear parameter model of (3.54) on the left, and the three nonlinear parameter model of (3.55) on the right, fitted to the S&P500 index between January 1985 and the crash of October 1987 using functional linear least squares as in (4.4). It can be quite clearly seen that by reducing the number of nonlinear parameters, the topology of the error function becomes smoother, and areas that give rise to large changes in the r.m.s for small changes in the parameter are reduced. However, significant sloppiness remains, especially on the surface the error function makes with t_c and β.

equation (3.54) to (3.55), and this is the functional form of the LPPL models that has been used in computation-tool development and data-fitting trials during the course of the research on which this thesis is based. Despite this reduction in sloppiness, one may remain unconvinced that there is any significance in the value of any particular parameter if one is focussed on finding the minimal r.m.s. error, and as such, this chapter focusses mainly on the method used most widely. At the end of this chapter, a novel approach is presented which is based on identifying the "gross" features of a time series, and thereby deriving a simple distribution of the non-linear parameters.

In the formulae for the various models in Table 3.1, the index $I(t)$ is a linear function of several of the parameters e.g. A, B, C_1 and C_2. These parameters are referred to as the *linear parameters* of the model. We call the other parameters, most particularly t_c, ω and β, *nonlinear parameters*. In common with the approach of Sornette and co-workers, for each choice of nonlinear parameters, one may use a standard least-squares algorithm to fit the remaining linear parameters, thereby reducing significantly the computational task involved.

4.1 Fitting with linear least squares and the taboo search algorithm

When one fits a function $f(x;\mathbf{p})$ with parameter vector $\mathbf{p} = (p_1, p_2, \ldots, p_m)$ to some observed data (x_i, y_i), where $i = 1, 2, \ldots, n$, one approach is to find the parameters $\mathbf{p}$ which minimise the sum of squared errors

$$\begin{aligned} e(\mathbf{p}) &= \sum_{i=1}^{n} (y_i - f(x_i))^2 \\ &= \sum_{i=1}^{n} y_i^2 - 2y_i f(x_i) + f(x_i)^2 \end{aligned} \tag{4.1}$$

such that

$$\frac{\partial e}{\partial p_j} = 2\sum_{i=1}^{N} \frac{\partial f(x_i)}{\partial p_j}(y_i - f(x_i)) = 0, \quad j = 1, 2, \ldots, m\,. \tag{4.2}$$

When the function, f, is composed of a linear combination of the parameters, the equation (4.2) may be solved analytically. Since each of the LPPL models can be expressed as

$$I(t) = A + Bf(x) + Cg(x) + Dh(x) \tag{4.3}$$

the sum of squared errors can be minimised by solving the system of linear equations (Filimonov & Sornette 2013)

$$\begin{pmatrix} n & \sum f(x_i) & \sum g(x_i) & \sum h(x_i) \\ \sum f(x_i) & \sum f(x_i)^2 & \sum f(x_i)g(x_i) & \sum f(x_i)h(x_i) \\ \sum g(x_i) & \sum f(x_i)g(x_i) & \sum g(x_i)^2 & \sum g(x_i)h(x_i) \\ \sum h(x_i) & \sum f(x_i)h(x_i) & \sum g(x_i)h(x_i) & \sum h(x_i)^2 \end{pmatrix} \begin{pmatrix} A \\ B \\ C \\ D \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum y_i f(x_i) \\ \sum y_i g(x_i) \\ \sum y_i h(x_i) \end{pmatrix} \tag{4.4}$$

Clearly, this is dependent on reasonable values for the nonlinear parameters having been previously chosen by some method. For the first-order LPPL model, these parameters are t_c, β, and ω.

One could use a nonlinear least-squares algorithm such as the Levenberg-Marquardt algorithm as described in the followings section, but with so many local minima (as can be seen from 4.1), it is important to scan the ranges of the parameter space to find reasonable starting parameter values. A good strategy may be to randomly select values for the nonlinear parameters, and use these choices as inputs the linear equation (4.3) such that the sum of squared errors can be minimised by solving (4.4). By compiling a list of the results, the parameter choice with the smallest error can be taken as the best fit. For what follows, this method is denoted as the base strategy. Clearly, there is no guarantee that this method will be successful in finding the global minimum, so perhaps a better strategy would be to construct a three-dimensional grid of nonlinear parameter values (in the case of the first-order model) and map the minima by iterating through the entire grid. However, this could be computationally very expensive if a very fine grid is constructed, and given the sloppiness of the model, one might expect that the grid will need be very fine indeed. The method that many

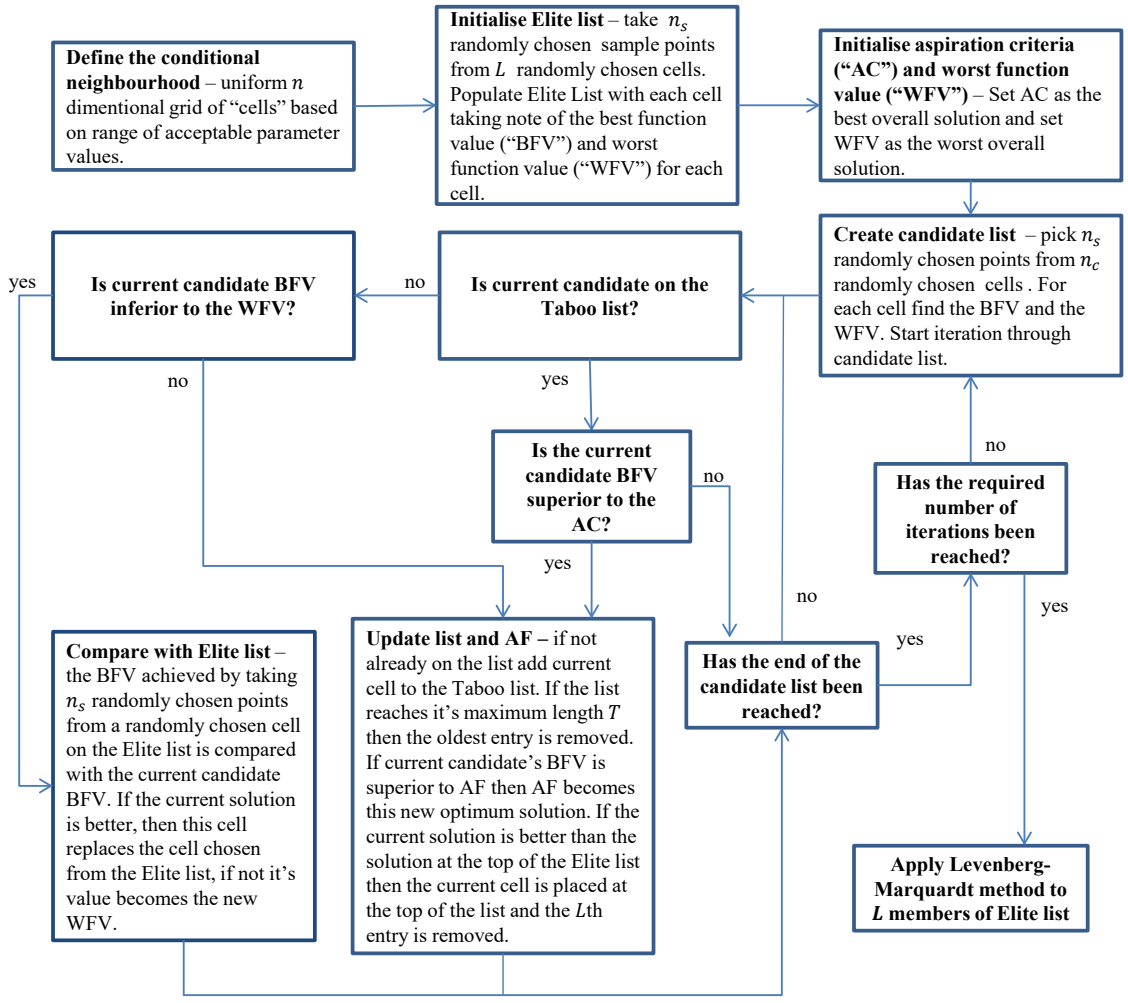


Figure 4.2: Taboo search flow chart

practitioners have used is a compromise between these to methods and is referred to as the taboo search (Cvijovic & Klinowski 1995).

The taboo search is a randomised hunt though the parameter space, but with a short term memory feature which steers the search away from previously visited fruitless areas to concentrate on unfamiliar or more successful regions. However, a region on the taboo list maybe redeemed if it can be shown that a parameter choice within this region can perform better than an "aspiration criterion". In the context of the problem at hand, it would seem reasonable to set this aspiration criterion to be the globally smallest r.m.s. error.

The flow chart of the taboo search methodology is shown in 4.2 and follows the algorithm suggested in Rajesh et al. (2000). The process starts by creating a coarse multidimensional grid spanning the parameter space. During the process two fixed-length lists are kept, the elite list, a fixed length record of those cells in the grid which have been found to harbour nonlinear parameters that give rise to the smallest r.m.s errors found so far, and the taboo list, a record of the most recently visited cells regardless of r.m.s. error performance.

The elite list is initialised by randomly selecting a number of cells in the grid, and computing the best function value (**BFV**) and worst function value (**WFV**) for each cell. The BFV and WFV for each cell are calculated by selecting random nonlinear parameter choices drawn from within the cell in question, applying (4.4), and computing the r.m.s. error arising from the fitted model. For each cell the the smallest r.m.s. error is denoted by the BFV and the largest by the WFV. The aspiration criterion (**AC**) is set as the overall smallest BFV achieved up to this point.

Once the elite list is initialised, a "candidate list" is drawn from randomly chosen cells in the grid, and each candidate is taken in turn. If the candidate cell is on the taboo list, then the only way the candidate can get on the elite list is if its BFV is superior to the the AC, otherwise it is ignored. However, if the candidate is not on the taboo list, the elite list is updated with candidate cells that perform better than cells currently on that list, as long as the candidate's BFV is not greater than (i.e. inferior to) the overall WFV seen so far. After

updating the elite list, the taboo list is updated accordingly.

Once the candidate list has been exhausted, a new candidate list is chosen and the process begins again. After sufficient iterations, the elite list is expected to contain the best possible solutions from a r.m.s. error perspective.

Once the process has finished, the non-linear parameter choices and the analytically found linear parameters are used as starting values for the Levenberg-Marquardt algorithm, which is described in the next section.

It should be noted that the taboo search has been used by many of the practitioners in this field, but its usefulness in fitting models of this type has not been borne out during the course of the research upon which this thesis is based. Indeed, the taboo search appears to have very little benefit over using a data set of randomly drawn parameters (as in the base strategy) to find starting values for the Levenberg-Marquardt algorithm. An example of the relative performance of each method is given in Fig. 4.3 where it can be seen that, for equal numbers of calculations, neither does the taboo search method settle on nonlinear parameter choices in tighter ranges, nor does it find lower r.m.s. errors. It could be conjectured that the reason for this is that 1) local minima of the error function are so closely spaced that a coarse-grained conditional neighbourhood grid is rendered useless, and/or 2) the set of solutions that are close to global minima is spread widely across the parameter space (see Fig. 4.4) in regions that are not related by any particular topology. Therefore, contrary to many researchers, the research described in the rest of this thesis does not employ this method, and instead favours the base strategy for finding nonlinear starting values.

4.2 The Levenberg-Marquardt algorithm

It is not possible to minimise the r.m.s. error for the LPPL models analytically, and one must use iterative processes to find the a minimum. Unfortunately, it is very difficult to judge, especially in the LPPL case, whether these iterative techniques have found a truly global minimum. In the case at hand, we have already described how the error function has a large number of very closely spaced min-

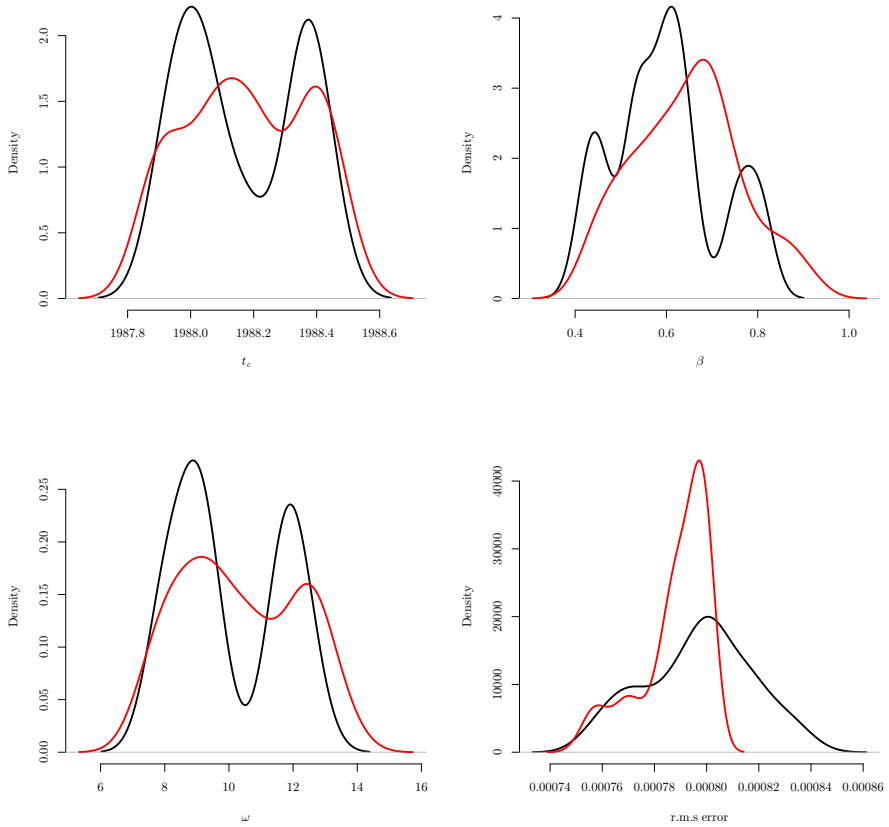


Figure 4.3: An example of the taboo search being used to find the nonlinear parameters for the first order LPPL model when fitted to the S&P500 index between January 1985 and the crash of October 1987. The parameter space was chosen such that $t_c \in [1987.85, 1988.50]$, $\beta \in [0, 1]$ and $\omega \in [0, 20]$. Both the elite list and taboo list were set at a maximum length of 30 entries. The parameter space was separated into a course grid of 100^3 boxes and the algorithm was iterated 10 times over a candidate list of length 100. The final distribution was found by sampling 10 random points in the top 10 cells of the elite list, whereupon the Levenberg-Marquardt algorithm was applied. The black lines show the nonlinear parameter and r.m.s error distributions resulting from applying the method. The red line shows the same distributions of the top 100 results emanating from randomly choosing the nonlinear parameter starting values from 6680 samples. This ensured the same number of calculations were performed as in the taboo search method. One may conclude, at least for this particular problem, the taboo search method may not give any particular advantage over the base method on in finding starting values for the LM algorithm.

ima, and, as such, the result of any nonlinear least squares method is dependent on the starting values chosen.

Once having scanned the parameter space for seemingly good parameter starting value candidates, many practitioners use the Levenberg-Marquardt (**LM**) algorithm (Gavin 2011) to iteratively seek a local minimum of the r.m.s. error between the LPPL model and the observed data. This method combines two distinct techniques, the gradient descent method and the Gauss-Newton method, so that the LM algorithm "graduates" between these two methods depending on whether or not the r.m.s. error is reduced by each iteration.

The gradient descent method simply takes the gradient of the error function with respect to the parameter vector $\mathbf{p}$ and delivers updates to the parameters choices which will reduce the error along this direction. Allowing $\mathbf{J}$ to be the Jacobian matrix where $J_{i,j} = \sum_i \sum_j \partial f(x_i)/\partial p_j$ and $f(x)$ to be the LPPL model to be fitted at points x_i, then equation (4.2) can be written as the partial derivative of the vector of errors, $\mathbf{e}$ with respect to the vector of parameters, $\mathbf{p}$, as follows (Gavin 2011)

$$\frac{\partial \mathbf{e}}{\partial \mathbf{p}} = -2(\mathbf{y} - \mathbf{f})^T \mathbf{J} \tag{4.5}$$

where $\mathbf{y}$ is the vector of observed points and $\mathbf{f}$ is the vector of the function f applied to the values of x_i with parameter vector $\mathbf{p}$. As such we can derive the direction, $\mathbf{h}_{sd}$, of the steepest descent to be $\mathbf{h}_{sd} = \mathbf{J}^T(\mathbf{y} - \mathbf{f})$.

On the other hand, the Gauss-Newton method assumes that that error function is approximately quadratic in its parameters close to the minimum, and one can write the error function (4.1) as $\mathbf{e} = \mathbf{y}^T\mathbf{y} - 2\mathbf{y}^T\mathbf{f} + \mathbf{f}^T\mathbf{f}$. Furthermore, the parameters in the function $\mathbf{f}(\mathbf{p})$ may be perturbed by an amount $\mathbf{h}_{gn}$, where the subscript denotes the perturbation is in respect of the Gauss Newton method, which then can be expanded in a Taylor series as

$$\mathbf{f}(\mathbf{p} + \mathbf{h}_{gn}) \approx \mathbf{f}(\mathbf{p}) + \mathbf{J}\mathbf{h}_{gn}\,. \tag{4.6}$$

This may be substituted into the expression for $\mathbf{e}$ as

$$\begin{aligned}\mathbf{e}(\mathbf{p}+\mathbf{h_{gn}}) &= \mathbf{y}^T\mathbf{y} - 2\mathbf{y}^T(\mathbf{f}+\mathbf{J}\mathbf{h}_{gn}) + (\mathbf{f}+\mathbf{J}\mathbf{h}_{gn})^T(\mathbf{f}+\mathbf{J}\mathbf{h}_{gn}) \\ &= \mathbf{y}^T\mathbf{y} - 2\mathbf{y}^T\mathbf{f} - 2\mathbf{y}^T\mathbf{J}\mathbf{h}_{gn} + \mathbf{f}^T\mathbf{f} + \mathbf{h}_{gn}^T\mathbf{J}^T\mathbf{J}\mathbf{h}_{gn} + 2\mathbf{f}^T\mathbf{J}\mathbf{h}_{gn} \\ &= \mathbf{y}^T\mathbf{y} - 2\mathbf{y}^T\mathbf{f} - 2(\mathbf{y}-\mathbf{f})^T\mathbf{J}\mathbf{h}_{gn} + \mathbf{f}^T\mathbf{f} + \mathbf{h}_{gn}^T\mathbf{J}^T\mathbf{J}\mathbf{h}_{gn}\,. \end{aligned} \tag{4.7}$$

Differentiating this expression with respect to the perturbation $\mathbf{h}_{gn}$ gives

$$\frac{\partial}{\partial \mathbf{h}_{gn}}\mathbf{e}(\mathbf{p}+\mathbf{h}_{gn}) = -2(\mathbf{y}-\mathbf{f})^T\mathbf{J} + 2\mathbf{h}_{gn}^T\mathbf{J}^T\mathbf{J} \tag{4.8}$$

which on setting $\partial\mathbf{e}/\partial\mathbf{h}_{gn} = 0$ we have the equation

$$\mathbf{J}^T\mathbf{J}\mathbf{h}_{gn} = \mathbf{J}^T(\mathbf{y}-\mathbf{f})\,. \tag{4.9}$$

Now, the LM algorithm gives the parameter update, $\mathbf{h}_{lm}$, such that

$$(\mathbf{J}^T\mathbf{J} + \lambda\mathbf{I})\mathbf{h}_{lm} = \mathbf{J}^T(\mathbf{y}-\mathbf{f}) \tag{4.10}$$

One can see that as the parameter λ becomes smaller, the resulting normalising equation tends to favour the Gauss-Newton method for parameter updating, whereas for larger λ, the parameter updating formula favours the steepest descent method.

In the course of this research, this algorithm has been implemented in a variety of different ways, but generally the following route was taken: initially a starting value for λ is chosen. Usually this value is large such that parameter updating is biased to steepest descent method. For most practical applications for the purposes of this research, the starting value of λ was 10^4. From here, the parameter update vector $\mathbf{h}_{lm}$ is found and the starting parameters are updated. Then a check is made to see if the error function has been reduced. If so, the new parameter values are accepted and the value of λ is reduced by a factor of 10, if not then the parameter update is ignored and the value of λ is increased by a factor of 10. The procedure is repeated until the change in the value of the

error function has less than some chosen threshold whilst keeping $\lambda < 10^{-8}$, or $\lambda > 10^{8}$.

One can question the suitability of the LM algorithm as it is applied to the current problem of fitting the LPPL models to observed data. It has been shown in Fig. 4.1 how model sloppiness may be reduced by eliminating one of the nonlinear parameters, but this does not necessarily alter the fact that r.m.s. errors of equal size may be found across parameter space, with large changes in some parameters having little effect, whist small changes in others having large effects. When the LM algorithm is applied to randomly chosen starting values of the nonlinear parameter space, along with those value of the linear parameters that satisfy the linear best fit equation (4.4), one can see the distribution of the set of derived local minima still exhibits similar sloppiness in relation to the nonlinear parameters associated with these local minima.

Fig. 4.4 shows an example of this problem. In accordance with the conclusion of the previous section, a large set of randomly generated starting values were taken, to which the best fitting linear parameters were found by application of (4.4). The LM algorithm was applied to this parameter set and the corresponding local minima derived by the algorithm were found. This gave a set of local minima where these local minima occur within the parameter space. One can see from the plot in Fig. 4.4 that there is not a unique global minimum, and by examining the how the best r.m.s. errors derived from the LM algorithm are distributed, it can be seen that although there are relationships between the nonlinear parameters, it is not possible to say that there is a particular nonlinear parameter choice that can be interpreted as being better than another.

The unsatisfactory nature of these results has led to some researchers in this field, and indeed the author of this thesis, to investigate methods that seek to describe the gross features of the observed data in more meaningful ways such that the nonlinear parameter choice is further limited. One of these methods which has been adopted as the favoured approach throughout this research is described in the next section.

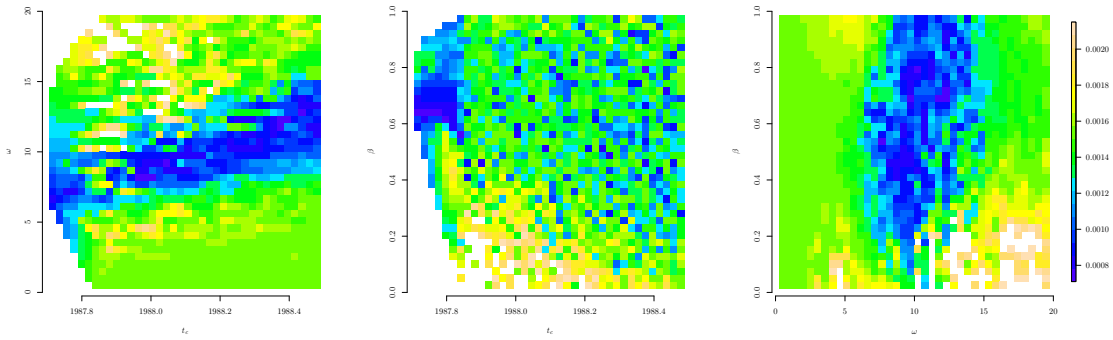


Figure 4.4: Nonlinear parameters of the first order LPPL model fitted to the S&P500 index between 1985 and the crash of October 1987. In fitting the model, 20,000 sets of randomly generated nonlinear parameters were taken (uniformly drawn from the parameter spaces in Fig. 4.2) and used as starting values in the LM algorithm. The plots above show the r.m.s. errors achieved (where smaller errors are indicated by the blue areas) for parameter pair combinations where the r.m.s error is within the right side 95% quantile, thus filtering high error outliers from the data set. As can be seen from these figures, although there are relationships between the nonlinear parameters, equally as good r.m.s. errors may be achieved with very different nonlinear parameter values.

4.3 Alternative approach

When comparing the hierarchy of LPPL models one may avoid the computational difficulties of the Levenberg-Marquardt algorithm by focussing on the interdependencies of a subset of nonlinear parameters. As in several works by D. Sornette and his co-workers (e.g. Sornette 2017) and Vandewalle, Boveroux, Minguet & Ausloos (1998), an approach is proposed that uses the gross features of the time series to obtain a relationship between the critical time t_c and the angular log-frequency ω.

Denoting by t_k the times at which the time series is a local maximum, as determined by eye and ignoring smaller scale oscillations, and considering two pairs of consecutive maximal points, (t_1, t_2) and (t_3, t_4), one can see that the angular frequency implied by the first pair, $\omega_{1,2}$ is given by

$$\omega_{1,2} = \frac{2\pi}{\log(t_2 - t_c) - \log(t_1 - t_c)} \tag{4.11}$$

and for the second pair

$$\omega_{3,4} = \frac{2\pi}{\log(t_4 - t_c) - \log(t_3 - t_c)} \tag{4.12}$$

Therefore, by assuming a constant angular log-frequency in the log-periodic oscillations, i.e. $\omega_{1,2} = \omega_{3,4}$, one may estimate the critical time using the cross-ratio condition $(t_c - t_1)/(t_c - t_2) = (t_c - t_3)/(t_c - t_4)$. This gives an estimate for the critical time t_F implied by the first-order model

$$t_F = \frac{c_2}{c_1}, \quad c_1 = t_1 + t_4 - t_2 - t_3, \quad c_2 = t_1 t_4 - t_2 t_3\,, \tag{4.13}$$

and it follows that an estimate for first-order angular log-frequency, ω_F, is given by

$$\omega_F = \frac{2\pi}{\log(t_F - t_1) - \log(t_F - t_2)}\,. \tag{4.14}$$

This method is illustrated for the S&P500 index in the period 1980–1988 in Fig. 4.5. If it is required for the first order-model to fit the angular log-frequencies in both the period between 1980–1984 and and the period between 1986–1987.5, thus "solving" the issue of the apparent frequency shift shown in Fig. 3.6, then, by (4.13), the consecutive peaks at 1980.91 and 1983.78, and later at 1986.5 and 1987.2, give a unique value for $t_F = 1988.30$. Having found this unique value of t_F, from (4.14) one may find that the unique value of ω is given by $\omega = 12.74$. This angular log-frequency is very close to the value found in Feigenbaum & Freund (1996) for the plot between 1980 and the crash of October 1987. This begs the question, what is the meaning of the apparent frequency shift that caused concern to Feigenbaum, and encouraged Sornette and his co-workers to develop the nonlinear correction to scaling given by the second-order model described in Chapter 3? Fig. 4.5 shows that in fact there is no frequency shift if one relaxes the condition on t_c to occur around the time of the crash itself. Indeed, there is nothing in the underlying theory that suggests that the critical time t_c should be coincident with the time of the crash, but rather it is the time at which a crash would become most likely.

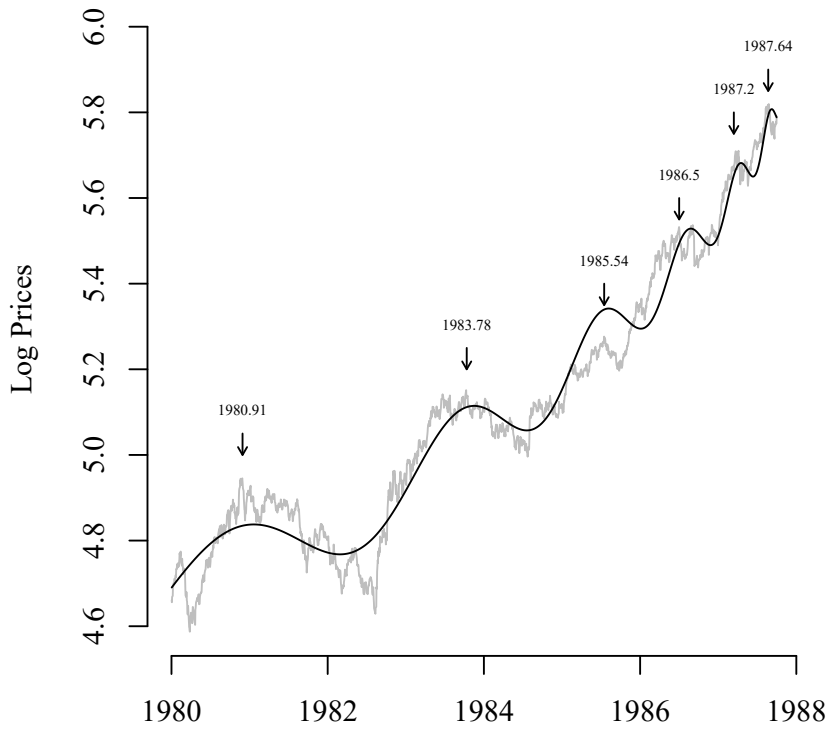


Figure 4.5: First-order fit of the S&P500 index 1980–1988. The times corresponding to peaks in the oscillations are marked by arrows. Assuming a constant angular log-frequency ω throughout the period, and requiring the model to fit the log-frequency of the peaks between 1980 and 1984 and between 1986 and 1987.50, the value $t_F = 1988.30$ is derived using the cross-ratio condition where $t_1 = 1980.91, t_2 = 1983.78, t_3 = 1986.50$, and $t_4 = 1987.2$. This value of t_F implies $\omega = 12.79$, and finally $\beta = 0.42$ is found using the LM algorithm to minimise the r.m.s. error. Sornette (2017) shows many more examples of such oscillations prior to markets crashes, for example, the crashes on the Hang Seng index in 1994 and 1997. However, the crash of October 1987 on the S&P500 index is the clearest example of this phenomenon.

In the case above, these two pairs of consecutive maximal points were chosen to consider a particular purpose of fitting two distinct parts of the time series in question. However, of course, there is no reason why any one pair of maximal points should be chosen over another. It is therefore worth analysing of each of these points in relation to each of the others. The results of this analysis can be seen in Table 4.1.

Table 4.1 shows the distribution of the values of t_F and ω_F that are derived from the cross-ratio condition of all combinations of pairs of consecutive peaks, and it can seen that although there is not a high density of data points, there is

t_1	t_2	t_3	t_4	t_F	ω_F	β_F	error
1980.91	1983.78	1983.78	1985.54	1988.33	12.85	0.42	0.0032
1980.91	1983.78	1985.54	1986.50	1987.87	11.81	0.50	0.0027
1980.91	1983.78	1986.50	1987.87	1988.30	12.79	0.42	0.0031
1980.91	1983.78	1987.20	1987.64	1988.34	12.87	0.41	0.0032
1983.78	1985.54	1985.54	1986.50	1987.65	10.37	0.55	0.0035
1983.78	1985.54	1986.50	1987.20	1988.27	12.72	0.42	0.0032
1983.78	1985.54	1987.20	1987.64	1988.34	12.88	0.42	0.0022
1985.54	1986.50	1986.50	1987.20	1989.08	19.89	0.31	0.0052
1985.54	1986.50	1987.20	1987.64	1988.60	16.72	0.31	0.0035
1986.50	1987.20	1987.20	1987.64	1988.39	13.53	0.41	0.0027

Table 4.1: The nonlinear parameter distribution given by adjacent peaks of the S&P500 index between 1980 and the crash of October 1987. The highlighted row contains the pairs used to produce the plot in Fig. 4.5.

the semblance of a somewhat symmetrical distribution of these values, to which one can find the corresponding minimum of the value for β_F, by using the LM algorithm as describe previously in this Chapter, and perhaps a most likely value (**MLV**) can be extracted from this sparse distribution of "optimum" parameter values.

However, a more robust method could be to use the data from these sparse distributions of t_F and ω_F as prior distributions in a Markov Chain Monte Carlo (**MCMC**) method of obtaining samples from the joint distribution of these parameters. MCMC methods such as the Metropolis–Hastings (Hastings 1970), and the Gibbs sampling (Geman & Geman 1984) algorithms, are implementations of Bayesian inferencing which allow sequences of samples to be indirectly collected where direct sampling of the underlying joint distribution of the parameters is either very difficult or impossible.

In order to do this, one must make assumptions concerning the shape and independence of these prior distributions, which, given the density of the data at hand, is only ever going to be a crude estimation. Therefore, for the purposes of simplicity, it has been assumed that these parameters are independent and approximately normally distributed with mean and standard deviation, such that

$$t_F \sim N(1988.32, 0.38), \quad \omega_F \sim N(13.64, 2.70) \tag{4.15}$$

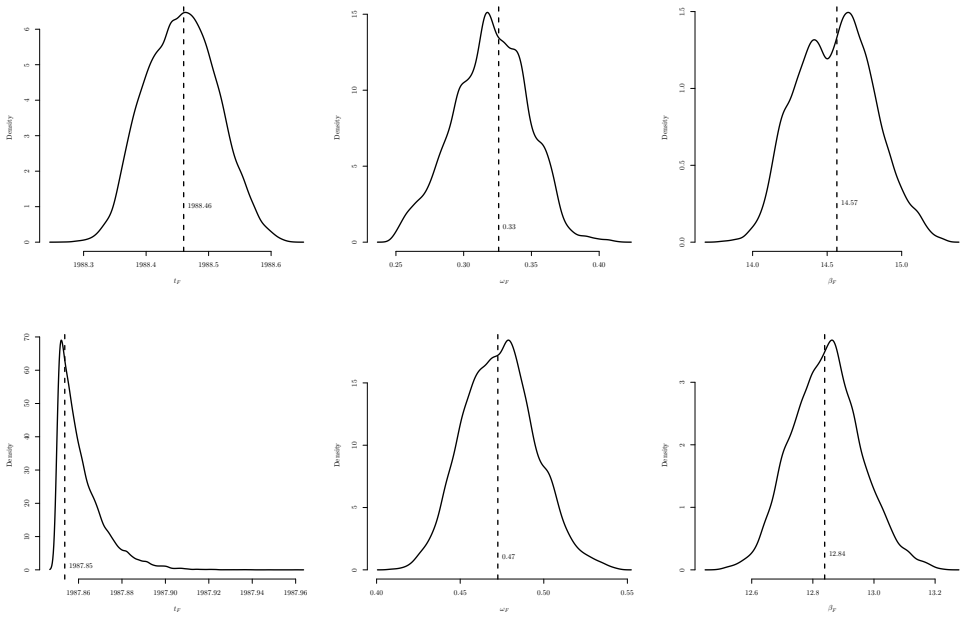


Figure 4.6: Posterior distributions of the nonlinear parameters of the first order model as fitted to the S&P500 index in the 8 year period prior to Black Monday using the priors given by the distributions suggested by Table 4.1 (top row), and those given by using uniform priors (bottom row) of the nonlinear parameters. As can be seen, using uniform priors the MLV of t_F is situated close to the time of the crash, which it has been demonstrated cannot be the case if the first order fit is to have a constant angular frequency. Each MCMC simulations used three chains with 1,000 burn-in iterations.

with β_F drawn from an uniform distribution between 0 and 1 such that $\beta_F \sim$ U(0,1).

These rough estimates of the prior distribution of t_F and ω_F result in an acceptable convergence (given by the Gelman-Rubin diagnostic) of the MCMC simulations, and it is shown that good convergence is also achieved when uniform priors are applied to these parameters (Fig. 4.6). However, in Fig. 4.7 it can be seen that the integrity of the gross feature fitting is lost when uninformed priors are used to find the nonlinear parameters with the Gibbs sampling method. It should be noted that the software used for these simulations was the `R2jags`[1] package as used in the statisitcal computing software `R`[2].

[1]https://cran.r-project.org/web/packages/R2jags/index.html
[2]https://www.r-project.org/

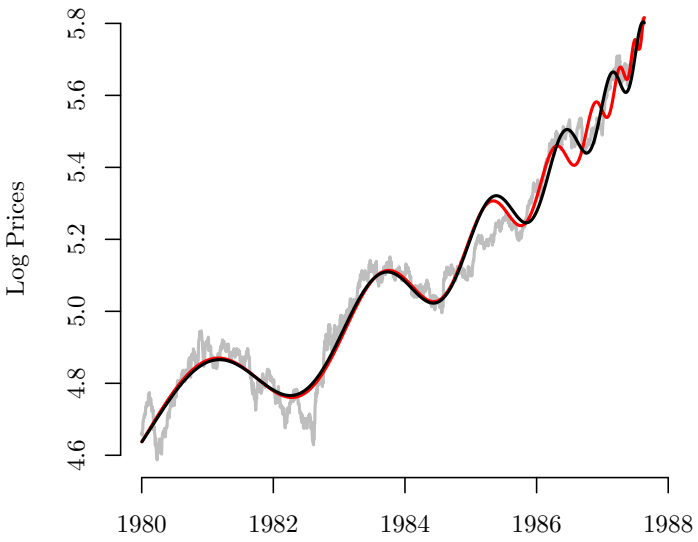


Figure 4.7: Plot of the first order fit of the S&P500 index using the MLV nonlinear parameters derived by the Gibbs sampling method using priors derived from the cross ratio criterion (black line). The red line shows the first order fit derived from using uninformed priors for t_F and ω_F. It can be seen this results in over oscillations as the crash approaches.

4.4 Summary

This chapter has shown the methods for fitting LPPL models to financial data that have been used by practitioners in this field, where focus has been given to the first order LPPL model. It has concluded by presenting a novel methodology using data derived from recognising that the gross features of the data can give important prior information about the distribution of the problematic nonlinear parameters. It is this method which is used for model fitting throughout what follows in this thesis.

During the investigation of fitting methods, it has also been shown that in the case of this first order model and the S&P500 index up to the Black Monday market crash, the value of t_F does not seem to be a good predictor of the time of the crash. However, no claim is being made that the critical time should indeed be the actual time of the crash, but the analysis presented here merely confirms that the first-order model cannot fit the gross features of both the long-term and the short-term S&P500 index time series over the 9 years prior to Black Monday with a constant angular log-frequency if one assumes the time of the crash and the critical time are coincident. In the next chapter, a revised model is presented which resolves this issue without resorting to the varying angular frequency of the higher order LPPL models.

Chapter 5

Logistic model for financial bubbles and anti-bubbles

In this chapter, a step back is taken from the evolution of the ideas developed over the past twenty-five years, and the initial renormalisation group formalism, from which the series of LPPL models have been derived, is revisited. In the spirit of the work of Curtright & Zachos (2011) in the physical context, a new series of models derived from the logistic differential equation is suggested, giving a non-linear perturbation of the standard linear transformation in the renormalisation group equation (Lynch & Mestel 2017).

5.1 Solutions to the renormalisation group equations and the Schröder equation

In the full renormalisation group equation given in (3.44), the non-singular element of $F(x)$ is characterised by some differentiable function $g(x)$, with $g(0) = 0$. For the purposes of this analysis and following most authors, it is considered that $g(x) \equiv 0$; for a discussion of the full equation and the significance of $g(x)$ see Gluzman & Sornette (2002). Therefore, it is taken that

$$F(x) = \frac{1}{\mu} F(\phi(x)) \,. \tag{5.1}$$

Note that (5.1) is a generalisation of the classical Schröder equation $\psi(\phi(x)) = \lambda\psi(x)$, where $\psi(x) = F(x)$. Indeed, for the case $\phi(x) = \lambda x + o(x)$ as $x \to 0$ and $\mu = \lambda \neq 0$, equation (5.1) reduces to $\lambda F(x) = F(\phi(x))$ so that F conjugates $\phi(x)$ to its linear part. For $\mu \neq \lambda$ the solution $F(x)$ of (5.1) has singular behaviour at $x = 0$.

Supposing that $\phi(x)$ is linear, such that one may write $\phi(x) = \lambda x$, (5.1) becomes

$$F(\lambda x) = \mu F(x)\,. \tag{5.2}$$

If one assumes that a solution to this equation is $F(x) = x^\beta$, then

$$(\lambda x)^\beta = \mu x^\beta \tag{5.3}$$

giving the condition $\beta = \log\mu/\log\lambda$. One may make the further assumption that the general solution to (5.1) is given by $F(x) = x^\beta G(f(x))$, where G and f are some, as yet unknown, functions. This being the case, it can be seen from (5.1) that

$$(\lambda x)^\beta G(f(\lambda x)) = \mu x^\beta G(f(x))\,, \tag{5.4}$$

such that clearly, $G(f(x)) = G(f(\lambda x))$. If one allows $f(x) = 2\pi\log x/\log(\lambda)$ then

$$\begin{aligned} G(f(\lambda x)) &= G\left(\frac{2\pi}{\log\lambda}\log(\lambda x)\right) \\ &= G\left(2\pi + \frac{2\pi}{\log\lambda}\log x\right) \end{aligned} \tag{5.5}$$

so that one may write the general solution to (5.1) as

$$F(x) = x^\beta G(\omega\log x)\,, \tag{5.6}$$

where $\beta = \log\mu/\log\lambda$, $\omega = 2\pi/\log\lambda$ and G is a periodic function in $\log x$. Therefore, $F(x)$ is also periodic in $\log(x)$ with period $\log\lambda$ and can be expressed as a Fourier series such that the truncated series can be written in the familiar

LPPL form

$$F(x) = x^{\beta} \left(B + C_1 \cos(\omega \log x) + C_2 \sin(\omega \log x) \right) . \tag{5.7}$$

Now, returning to the general case of nonlinear $\phi(x)$, one can solve (3.44) by making a change of x-coordinate. Firstly, it is assumed there is a solution $\psi(x)$ around $x = 0$ of the Schröder equation $\psi(\phi(x)) = \lambda\psi(x)$, satisfying $\psi(0) = 0$, and for convenience, the normalisation $\psi'(0) = 1$. Therefore, $\psi(x)$ conjugates $\phi(x)$ to its linearisation λx around $x = 0$. When $\psi(x)$ is known, it is straightforward to reduce (3.44) to the linear case by making the change of coordinate[1] $x \to \psi(x)$, so that writing $F(x) = \tilde{F}(\psi(x))$ gives

$$\tilde{F}(\psi(x)) = \frac{1}{\mu}\tilde{F}(\lambda\psi(x)) = \frac{1}{\mu}\tilde{F}(\psi(\phi(x))) , \quad \psi(x) > 0 . \tag{5.8}$$

Then, from (5.6) we obtain the general solution for $F(x)$ of (3.44), at least for $x > 0$ sufficiently small, in the form:

$$F(x) = \tilde{F}(\psi(x)) = \psi(x)^{\beta} G(\omega \log \psi(x)) , \quad x > 0 , \tag{5.9}$$

where $\beta = \log \mu / \log \lambda$ and $\omega = 2\pi / \log \lambda$. Note that (5.9) reduces to (5.6) in the case $\phi(x) = \lambda x$ (for which $\psi(x) = x$). Therefore, in general, finding other solutions to the Schröder equation will lead to other, nonlinear perturbations of the linear renormalisation group equation.

However, in general the Schröder equation cannot be solved in closed form, but when $\phi(x)$ is the time-1 map of an autonomous differential equation, the conjugating function $\psi(x)$ may be readily found. Consider the differential equation

$$\frac{dx}{dt} = \rho h(x) , \quad x(0) = x_0 , \quad h(x) = x + O(x^2) . \tag{5.10}$$

The solution $x(t)$ of (5.10) is obtained implicitly by separation of variables in

[1] in the process of reviewing Lynch & Mestel (2017) D. Sornette kindly pointed out that this change of coordinate could be interpreted as a nonlinear map from calendar time to an "investor time" in line with the concept of subordination. See Geman & Ané (1996) and Mandelbrot et al. (1997).

terms of the function

$$H(x(t)) \stackrel{\text{def}}{=} \int \frac{dx}{h(x)} = \rho t + C\,, \tag{5.11}$$

where C is a constant of integration, which gives $H(x(t)) - H(x_0) = \rho t$ when $x(0) = x_0$. Exponentiating this expression gives

$$\exp[H(x(t))] = \exp[H(x_0)]e^{\rho t}\,. \tag{5.12}$$

Now, defining $\phi(x_0) = x(1)$, $\psi(x_0) = \exp[H(x_0)]$ and $\lambda = e^{\rho}$, the functional equation $\psi(\phi(x_0)) = \lambda\psi(x_0)$, is obtained as required. Furthermore, $H(x) = \log x + C + o(1)$ as $x \to 0+$, where C is an arbitrary constant of integration. Setting $C = 0$, gives, on taking the limit $x \to 0+$, the normalisations $\psi(0) = 0$ and $\psi'(0) = 1$.

5.2 Logistic differential equation

Consider the special case of the logistic differential equation.

$$\frac{dx}{dt} = \rho x(1 + \nu x)\,, \quad x(0) = x_0\,, \quad \rho = \log\lambda\,, \tag{5.13}$$

Separating the variables gives the integral

$$\begin{aligned}\rho t + C &= \int \frac{dx}{x(1+\nu x)} \\ &= \int \left(\frac{1}{x} - \frac{\nu}{1+\nu x}\right) dx \\ &= \log\left(\frac{x}{1+\nu x}\right)\,. \end{aligned} \tag{5.14}$$

Since $x(0) = x_0$, and setting $C = 0$, the solution

$$x(t) = \frac{e^{\rho t}x_0}{(1 + \nu x_0(1 - e^{\rho t}))} \tag{5.15}$$

is given and fractional-linear time-1 map where $\phi(x_0) = x(1)$

$$\phi(x_0) = \frac{\lambda x_0}{1 + \nu x_0(1-\lambda)}\,, \tag{5.16}$$

which, on replacing x_0 by x, gives a nonlinear perturbation of $\phi(x) = \lambda x$, parametrised by ν. The conjugating function $\psi(x)$ may also be calculated explicitly from $\psi(x_0) = exp[H(x_0)]$:

$$\psi(x) = \frac{x}{1 + \nu x}\,. \tag{5.17}$$

The behaviour of the solution is dependent on the sign of ν. There is a finite-time singularity in the solution of (5.13) when $\nu > 0$, and the map $\phi(x)$ has a singularity at $x = (\nu(\lambda-1))^{-1}$, but the function $\psi(x)$ is well defined for all $x \geq 0$, and saturates at ν^{-1} as $x \to \infty$. However, for $\nu < 0$, $\phi(x)$ is finite for $x > 0$, but $\psi(x)$ has a singularity at $x = -\nu^{-1}$, and therefore $F(x)$ has a singularity for $x > 0$ as well as at $x = 0$.

By writing $\nu = 1/\Delta_t$, and remembering that in this context, generally $x = |t_c - t|$, the general solution may be written as

$$\tilde{F}(x) = \left(\frac{|t_c - t|}{1 + \frac{|t_c - t|}{\Delta_t}}\right)^{\beta} G\left(\omega \log\left(\frac{|t_c - t|}{1 + \frac{|t_c - t|}{\Delta_t}}\right)\right) \tag{5.18}$$

Since it is known that G is a periodic function in $\log \psi(x)$ with period $\log \lambda$, following (5.7), this nonlinear perturbation, which has been called the *logistic model* (Lynch & Mestel 2017), may be cast in a form comparable to those in Table 3.1:

$$\begin{aligned} I(t) &= A + \frac{|t_c - t|^{\beta}}{\left(1 + \frac{|t_c - t|}{\Delta_t}\right)^{\beta}} \left[B + C_1 \cos\theta(t) + C_2 \sin\theta(t)\right]\,, \\ \theta(t) &= \omega \log|t_c - t| - \omega \log\left(1 + \frac{|t_c - t|}{\Delta_t}\right)\,. \end{aligned} \tag{5.19}$$

Whilst the second-order and logistic models coincide with the first-order model for small $|t_c - t|$, the properties of the logistic model (5.19) for large $|t_c - t|$ are distinct

from those of (2.4). For $\Delta_t > 0$, both models lead to saturation of the amplitude with limit Δ_t^β. However, model (2.4) exhibits a frequency shift $\omega \to \omega + \Delta_\omega$ in the log-periodic oscillations (since $\theta(t) \sim (\omega + \Delta_\omega) \log |t_c - t| - \Delta_w \log \Delta_t$ for large $|t_c - t|$), whilst in model (5.19) $\theta(t) \to \omega \log \Delta_t$ as $|t_c - t| \to \infty$ and the log-periodic oscillations die away. This difference can be used to determine which model is of more practical usefulness in analysing various asset bubbles and anti-bubbles. For $\Delta_t < 0$, the logistic model has a secondary critical point at $|t_c - t| = |\Delta_t|$.

This means that the log-periodic oscillations leading up to this secondary critical point grow without bound. Throughout the literature on this topic, the aim has always been to consider cases only where $F(x)$ remains bounded, since singular behaviour is not seen in real world observations. This suggests one should neglect cases where $\Delta_t < 0$. However, as we will see later, allowing this situation can lead to some interesting results, and the model can nevertheless be used to obtain remarkable fits in some cases, implying both beginning and end points to regions of herding behaviour. It should be noted that the LPPL and logistic models are applicable only during these periods of cooperative behaviour and are not valid outside the bounds of $t > t_c$ for a bubble and $t < t_c$ for an anti-bubble. Moreover, for $\Delta_t < 0$, the logistic model is additionally not valid outside the bounds $t > t_c + \Delta_t$ and $t < t_c + |\Delta_t|$ for bubble and anti-bubble respectively. Applications of the logistic model are described in §5.4

5.3 Perturbation of the logistic differential equation

By taking higher-order non-linearities in the differential equation (5.13), it is possible to build a hierarchy of models as in the LPPL approach. For example, taking a perturbation of the logistic differential equation

$$\frac{dx}{dt} = \rho x(1 + \nu x)(1 + \sigma x)\,, \quad |\sigma| < |\nu|\,, \tag{5.20}$$

similarly results in (5.9) with

$$\psi(x) = \frac{x(1+\sigma x)^{\frac{\sigma}{\nu-\sigma}}}{(1+\nu x)^{\frac{\nu}{\nu-\sigma}}} \,. \tag{5.21}$$

Writing $\nu = 1/\Delta_t$ and $\sigma = \epsilon/\Delta_t$, gives

$$\psi(x) = \frac{x\left(1+\frac{\epsilon x}{\Delta_t}\right)^{\frac{\epsilon}{1-\epsilon}}}{\left(1+\frac{x}{\Delta_t}\right)^{\frac{1}{1-\epsilon}}} \tag{5.22}$$

leading to the so called *modified logistic model* (Lynch & Mestel 2017):

$$I(t) = A + \frac{|t_c - t|\left(1+\frac{\epsilon|t_c-t|}{\Delta_t}\right)^{\epsilon/(1-\epsilon)}}{\left(1+\frac{|t_c-t|}{\Delta_t}\right)^{1/(1-\epsilon)}}\left[B + C_1\cos\theta(t) + C_2\sin\theta(t)\right] \,, \tag{5.23}$$

$$\theta(t) = \omega\log|t_c - t| - \frac{1}{1-\epsilon}\left(1+\frac{|t_c-t|}{\Delta_t}\right) + \frac{\epsilon}{1-\epsilon}\left(1+\frac{\epsilon|t_c-t|}{\Delta_t}\right) \,. \tag{5.24}$$

Although not further pursued for the purposes of this thesis, one should note that it is straightforward to adapt the techniques described in the remainder of this chapter to fit the modified logistic model (and other models obtained from other perturbations of the logistic differential equation) to financial time series displaying the characteristics of bubbles and anti-bubbles.

5.4 Comparison of model fitting for S&P 500 and Nikkei 225 indices

To investigate the performance of the logistic model in comparison to the established LPPL model hierarchy, it is useful to study examples of markets which have exhibited bubble or anti-bubble behaviour that have been extensively researched during the development of the LPPL field of study. In this way, one can establish a base line of model performance for models of this type against which the logistic model may be measured. As has been mentioned throughout this thesis, one market that has been thoroughly examined, and is ubiquitous in LPPL

literature, is the S&P500 index during the long bull market in the 1980s leading to the crash of October 1987. The performance of the LPPL model hierarchy for this asset bubble time-series is well documented particularly in the large body of research by Didier Sornette and his co-workers. This historical work gives a firm base to be used to gauge the comparative performance of the logistic model.

A well studied example of an anti-bubble is given by the period 1990–2003 during which the Japanese stock market experienced a long bear run following the Japanese asset price bubble of the late 1980s. Again, there is good historical precedent of analysis of this time-series in respect of the LPPL hierarchy, and is useful to see in this example the LPPL analysis extended to the third order model. Interesting literature is available based on the progression of the Nikkei 225 index during this period (Johansen & Sornette 1999b) and in particular the *ex ante* predictive qualities of the third order LPPL model.

The following sections describe in detail how the logistic model performs in comparison to the LPPL hierarchy for these two examples, showing how this non-linear perturbation of the linear model can out perform the established hierarchy in a a variety of ways.

5.4.1 Example 1: S&P 500 bubble from 1980 to October 1987

Firstly, turning to the long bull market in the 1980s preceding Black Monday, the method described in §4.3 is extended to take the additional nonlinear parameter, Δ_t, of the logistic model into consideration, and a comparison with the first and second order LPPL model is made with interesting results concerning the apparent frequency shifting of this bubble regime.

Frequency shifting in S&P500 and the logistic model

As has previously been discussed, fitting these nonlinear models to the observed data is challenging, and the term "best fit" in the context of minimising a measure of the fits r.m.s. error does not lead to a meaningful set of parameters, i.e. fitting

models of this type or inherently "sloppy".

It has been explained earlier that the second-order LPPL model was developed to explain the apparent frequency shift in the log-periodic oscillations of the S&P 500 index observed by Feigenbaum & Freund. However, is has been seen in §4.3 that by moving the critical time out further into 1988, the first-order model can fit the long-term data very well, and there is non need to assume that the observed data exhibits any frequency shifting at all. Nevertheless, if it required that the fitted critical time t_c is close to the actual time of the crash (and there is no reason one should make this restriction), the logistic model's nonlinear perturbation of the first-order LPPL model may also be of some assistance in this particular case.

In fitting the logistic model, a similar method to that presented in §4.3 can be used to obtain an estimate for t_L, the critical time associated with the logistic model. Using the times t_1, ..., t_4, as before, the relation for two consecutive times t_k and t_{k+1}, in the case of the logistic model since the frequency of oscillation of this model, ω_L is given by

$$\omega_L = \frac{2\pi}{\log\left(\frac{t_L - t_k}{1+\frac{t_L-t_k}{\Delta_t}}\right) - \log\left(\frac{t_L - t_{k+1}}{1+\frac{t_L-t_{k+1}}{\Delta_t}}\right)} \tag{5.25}$$

which gives, on exponentiating and equating for the cases consecutive peaks such that circumstances where $k = 1$ and $k = 3$ are considered,

$$\left(\frac{t_L - t_1}{t_L - t_2}\right)\left(\frac{\Delta_t + t_L - t_2}{\Delta_t + t_L - t_1}\right) = \left(\frac{t_L - t_3}{t_L - t_4}\right)\left(\frac{\Delta_t + t_L - t_4}{\Delta_t + t_L - t_2}\right) \tag{5.26}$$

On expanding, this gives the quadratic equation in t_L:

$$-c_1\Delta_t t_L^2 + (-c_1\Delta_t^2 + 2c_2\Delta_t)t_L + c_2\Delta_t^2 + c_3\Delta_t = 0 \tag{5.27}$$

where, as before $c_1 = t_1 + t_4 - t_2 - t_3$, $c_2 = t_1t_4 - t_2t_3$, and $c_3 = t_1t_2t_3 - t_1t_2t_4 -$

$t_1t_3t_4 + t_2t_3t_4$. The solution to this quadratic equation can be written as

$$t_L = -\frac{1}{2c_1\Delta_t}\left(c_1\Delta_t^2 - 2c_2\Delta_t \pm \sqrt{(-c_1\Delta_t^2 + 2c_2\Delta_t)^2 - 4(-c_1\Delta_t)(c_2\Delta_t^2 + c_3\Delta_t)}\right)$$
$$= \frac{c_2}{c_1} - \frac{\Delta_t}{2} \pm \sqrt{\frac{\Delta_t^2}{4} + \frac{c_2^2}{c_1^2} + \frac{c_1c_3}{c_1^2}} \tag{5.28}$$

Since $t_F = c_2/c_1$, a relationship can be found in terms of the the control parameter, Δ_t and the first-order estimated critical time, t_F, and is given by

$$t_L = t_F - \frac{\Delta_t}{2} \pm \sqrt{\frac{\Delta_t^2}{4} + \kappa} \tag{5.29}$$

where $\kappa = (c_2^2 + c_1c_3)/c_1^2 > 0$.

If one takes the positive root, then, as $\Delta_t \to \infty$, the critical time $t_L \to t_F$ from above, and, as $\Delta_t \to -\infty$, $t_L \to \infty$. This means that, by taking this positive root, it is impossible to find a value for $t_L < t_F$. On the other hand, if one takes the negative root, then, as $\Delta_t \to \infty$, the critical time $t_L \to -\infty$, but as $\Delta_t \to -\infty$, $t_L \to t_F$ from below. Since by taking the positive root, one cannot find a critical time closer to the time of the actual crash than t_F, and, if it is important that t_L is close to the actual crash time, we must take the negative root.

The equation (5.29) gives a relation between t_L and Δ_t. In the case of the S&P500 index between 1980–1987.8 we further assume, following Feigenbaum & Freund (1996), that $t_L = 1987.8$, and is coincident to the time of the crash. This implies $\Delta_t = -15.70$ from (5.29) with $t_1 = 1980.91, t_2 = 1983.78, t_3 = 1986.5$, and $t_4 = 1987.2$. Then from (5.25) one can calculate that $\omega = 7.65$.

The fit of this logistic model is shown in Fig. 5.1. Although to the eye the fit is not as good as the fit that can be achieved with the first-order model with $t_F = 1988.30$, the logistic model with $t_L = 1987.80$ does not show the over-oscillation of the first-order model with $t_F = 1987.80$. However, the logistic model does not perform as well as the second-order LPPL model of Sornette & Johansen (1997). Nevertheless, if one relaxes the requirement that the critical time is close to the observed crash, then the S&P500 series with the apparent frequency shifting observed by Feigenbaum and Freund can be well described by

a constant angular log-frequency with the nonlinear perturbation given by the logistic model. One should note, in fact, that the logistic model gives an equally good account of the observed data as the second-order model if one allows the value of t_c to vary away from the actual crash date. To match the second-order model in terms of r.m.s. errors, the value of t_c for the logistic model is approximately 1988, much close to the observed time of the crash than could be achieved by a frequency matching first order model.

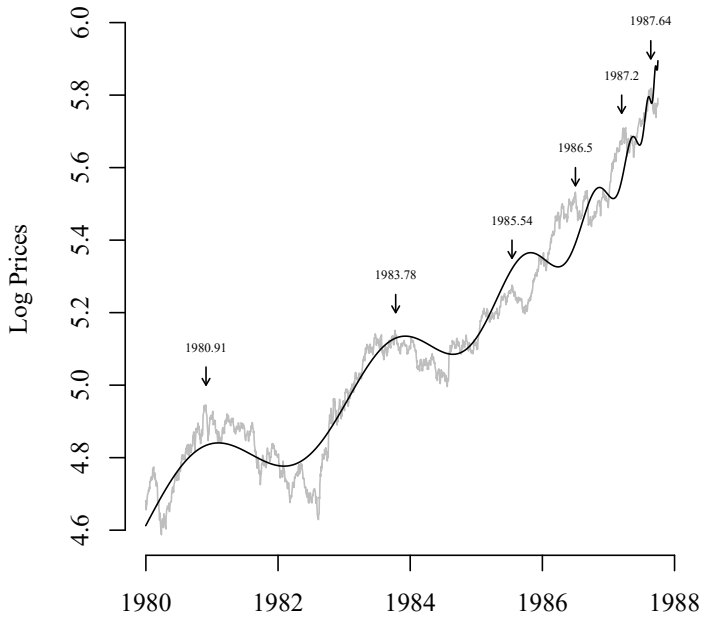


Figure 5.1: Fit of the logistic model to S&P500 1980–1988, with peaks marked by arrows. The parameter Δ_t is calculated from (5.27) the logistic model assuming a constant angular log-frequency ω and $t_c = 1987.8$, which gives $\Delta_t = -15.7$, and $\omega = 7.65$. The logistic model picks out the main features more accurately than the first-order model when one fixes the critical time to that of the crash itself, but not as well as the second-order model with the additional parameter, Δ_ω.

5.4.2 Example 2: Nikkei 225 anti-bubble from 1990 – 2003

The second case study is the Nikkei 225 index bear market during the period 1990–2003. In the following analysis time-series data between the known start and end dates of the anti-bubble is considered.

For anti-bubbles, one expects the critical time, t_c to be located prior to the

formation of the herding phenomenon which leads to a faster than exponential decline in asset prices. Therefore, each of the equations describing the LPPL hierarchy or logistic models must be rewritten to such that $t_c - t$ is replaced with $t - t_c$. Rewriting (5.26) as

$$\left(\frac{t_2 - t_L}{t_1 - t_L}\right)\left(\frac{\Delta_t + t_1 - t_L}{\Delta_t + t_2 - t_L}\right) = \left(\frac{t_4 - t_L}{t_3 - t_L}\right)\left(\frac{\Delta_t + t_3 - t_L}{\Delta_t + t_4 - t_L}\right), \tag{5.30}$$

one obtains the relationship between t_L and Δ_t for the logistic model of an anti-bubble regime:

$$\Delta_t = \frac{c_1 t_L^2 - 2c_2 t_L - c_3}{c_1 t_L - c_2} \tag{5.31}$$

where, as before,

$$c_1 = t_1 + t_4 - t_2 - t_3 \tag{5.32}$$

$$c_2 = t_1 t_4 - t_2 t_3 \tag{5.33}$$

$$c_3 = t_1 t_2 t_3 - t_1 t_2 t_4 - t_1 t_3 t_4 + t_2 t_3 t_4 \tag{5.34}$$

and the angular log-frequency is given by

$$\omega = \frac{2\pi}{\log\left(\frac{t_{k+1} - t_L}{1 + \frac{t_{k+1} - t_L}{\Delta_t}}\right) - \log\left(\frac{t_k - t_L}{1 + \frac{t_k - t_L}{\Delta_t}}\right)}. \tag{5.35}$$

As with the previous example, the gross features of the Nikkei 225 index between the origin of the anti-bubble at the beginning of 1990s to the end of the long down-turn in 2003 have been identified. However, in this example the peaks seem to be less well defined than that of the S&P 500 index between 1980–1988. Therefore, in contrast, the troughs (which are more well defined) have marked the periodicity.

A value for Δ_t can be determined using (5.31) by taking two pairs of consecutive troughs, and specifying the value of t_L. In an anti-bubble regime one expects the critical time to occur before the start of the regime. In this particular time series there is a sharp peak in the observations just before the beginning of the 1990s, so, in this case, the decision has been made that, first, it is important to

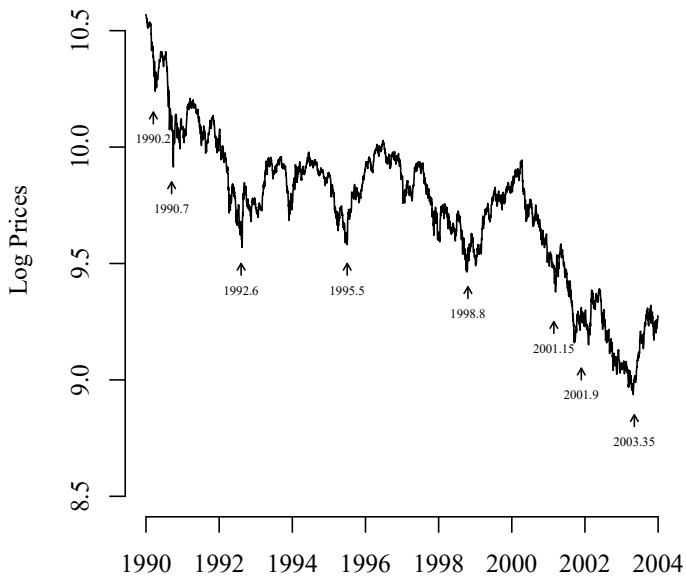


Figure 5.2: The peaks of the Nikkei 225 index are not very well defined so in this case the troughs are used as markers for the periodicity. As one can see, a marker at $t = 2001.15$ has been included but not so for the intermediate trough around $t = 1994$. The reason for this is that the latter dip seems not be in keeping with a general downward trend. Clearly this is a largely subjective judgement.

set the critical time at the point where the anti-bubble begins, and, second, that this beginning is at $t_L = 1989.95$.

As can be seen from Fig. 5.2, there are many pairs of consecutive troughs that can be chosen, and in contrast to the S&P 500 example where an attempt was made to correct the over-oscillations of the first-order model in a particular period, in the Nikkei 225 problem there are no *a priori* reasons to choose one pair over any other.

Therefore, in a similar way to that described in §4.3 values of Δ_t and the corresponding values for ω may be found using (5.31), and all combinations of consecutive pairs of troughs. The results of these calculations can be seen in Table 5.1. Apart from a few outlying data points (corresponding to pairs of troughs at the beginning and the end of the time period), the values of Δ_t calculated from picking out these troughs by eye show remarkable consistency over a very long time span. From this data, one is able to derive a limited distribution of values

t_1	t_2	t_3	t_4	ω	Δ_t
1990.20	1990.70	1990.70	1992.60	6.12	6.38
1990.20	1990.70	1992.60	1995.50	5.49	−11.35
1990.20	1990.70	1995.50	1998.80	5.51	−12.28
1990.20	1990.70	1998.80	2001.15	5.52	−12.80
1990.20	1990.70	2001.15	2001.90	5.51	−12.33
1990.20	1990.70	2001.90	2003.35	5.54	−14.21
1990.70	1992.60	1992.60	1995.50	3.98	−7.76
1990.70	1992.60	1995.50	1998.80	4.27	−10.75
1990.70	1992.60	1998.80	2001.15	4.35	−12.20
1990.70	1992.60	2001.15	2001.90	4.35	−12.20
1990.70	1992.60	2001.90	2003.35	4.43	−13.94
1992.60	1995.50	1995.50	1998.80	5.84	−12.78
1992.60	1995.50	1998.80	2001.15	5.89	−13.01
1992.60	1995.50	2001.15	2001.90	5.74	−12.37
1992.60	1995.50	2001.90	2003.35	6.16	−14.37
1995.50	1998.80	1998.80	2001.15	6.03	−13.09
1995.50	1998.80	2001.15	2001.90	5.55	−12.34
1995.50	1998.80	2001.90	2003.35	6.81	−14.56
1998.80	2001.15	2001.15	2001.90	4.35	−12.20
1998.80	2001.15	2001.90	2003.35	9.07	−15.25
2001.15	2001.90	2001.90	2003.35	-13.09	−10.16

Table 5.1: Implied values for ω and Δ_t from pairs of consecutive troughs.

for Δ_t and ω by including only the data from Table 5.1 corresponding to times within the data set being analysed.

In an anti-bubble regime, if the value of t_c signals the beginning of the anti-bubble, then the singularity at $t = t_c - \Delta_t$ signals its end. Therefore, one must also exclude values of Δ_t that imply an end to the anti-bubble prior to the end of the data set. If one accepts this argument, one must also exclude any positive value of Δ_t which implies the end of the anti-bubble being located before its start. However, logistic anti-bubble models with $\Delta_t > 0$ do not exhibit finite-time singularities but rather saturate as $t \to \infty$, which does not describe well the long-term progression of the Nikkei 225 index from 1990–2003.

Fitting the logistic model with Markov Chain Monte Carlo (MCMC)

In the logistic model there are four non-linear parameters to estimate, namely t_c, β, ω and Δ_t. For this example the critical time is fixed with $t_c = 1989.95$ and one may use the information about the distribution of both ω and Δ_t derived from Table 5.1. In the absence of any other prior information concerning the "real" distributions of ω and Δ_t, it seems reasonable to assume they are normally distributed, with some mean and standard deviation. Since Table 5.1 gives only a very limited sample size, the generalised Student's t-distribution, $t(\mu, \sigma^2, \nu)$, where

$$p(x|\nu, \mu, \sigma) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu\sigma^2}}\left(1 + \frac{1}{\nu}\frac{(x-\mu)^2}{\sigma^2}\right)^{\frac{\nu+1}{2}} \tag{5.36}$$

and μ is the mean and σ^2 is approximated as the variance, gives a reasonable prior for the purposes of the analysis which follows. We also know that $0 < \beta < 1$ so that the model remains bounded at t_c and is singular in its first derivative at t_c, hence we choose prior distributions for these parameters:

$$\beta \sim U(0,1), \quad \omega \sim t(\bar{\omega}, s_\omega, \nu), \quad \Delta_t \sim t(\bar{\Delta}_t, s_{\Delta_t}, \nu)\,, \tag{5.37}$$

where the means $\bar{\omega}$ and $\bar{\Delta}_t$, and the sample variances $s^2_{\Delta_t}$ and s^2_ω are calculated directly from Table 5.1 with respect to the data set in question making the above exclusions, and the degrees of freedom, ν, is the number of observations minus 1.

t_1	t_2	t_3	t_4	ω	Δ_t
1990.20	1990.70	1992.60	1995.50	5.49	−11.35
1990.70	1992.60	1992.60	1995.50	3.98	−7.76

Table 5.2: Implied values for ω and Δ_t from pairs of consecutive troughs for the period between 1990–1995.5 excluding positive values of Δ_t.

From here, the Gibbs sampling variation of MCMC simulation (using the `R2jags` software implemented in `R`) is used to generate posterior distributions for each of the non-fixed parameters (including the linear parameters), noting that in practice one needs to truncate the distribution for Δ_t such that a singularity cannot occur within the time span of the data series, and similarly the distribution for ω must be non-negative.

As an illustration, two examples from Johansen & Sornette (1999b) are taken, where the authors have fitted the second-order LPPL model to the period between 1990–1995.5 and the third-order model to the period between 1990–1999. These fits are compared with the results obtained from fitting the logistic model using the above method.

First, the trough pairs that are not suitable for this particular data set are excluded. As can be seen from Table 5.2 there are only two entries that are suitable giving $\bar{\Delta}_t = -9.55$ and $s^2_{\Delta_t} = 6.46$, and $\bar{\omega} = 4.73$ and $s^2_{\omega} = 1.13$. Then, as can be seen from Fig. 5.3, using the prior distributions in (5.37), the MCMC simulation produces approximately normal posterior distributions for the parameters ω and β, and a skew-normal distribution for Δ_t as follows

$$\beta \sim N(0.02, 0.04), \quad \omega \sim N(5.47, 0.10), \quad \Delta_t \sim SN(-35.80, 8.29, 0.50)\,. \tag{5.38}$$

Here the skew normal distribution, $SN(\zeta, \omega, \alpha)$ is parameterised by the location, ζ, the scale, ω, and the shape α. Using the mode of each of the distributions in (5.38) and the fixed value for t_c we can find the linear parameters analytically by minimising the r.m.s. errors. It is interesting to note that 95% confidence interval for Δ_t is between -61.97 and -22.65. Following our earlier argument, this puts the date for the end of the crash between 2012.65 and 2051.97. However,

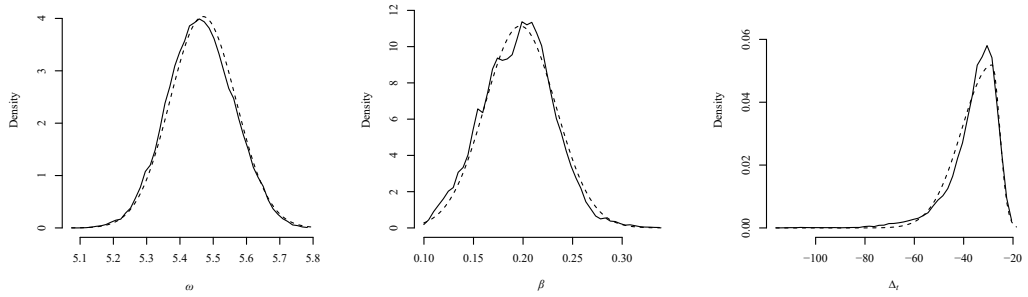


Figure 5.3: Density plots of posterior distributions (with 10,000 MCMC iterations) for the logistic model fitted between 1990–1995.5 with approximate distributions (dotted lines), normal for ω and β and skew-normal for Δ_t

one can see that the anti-bubble ended in the first half of 2003 and the calculated value, i.e. the mode of the posterior distribution, is $\Delta_t = -20.31$ which is not representative of the end date of the anti-bubble.

The results of this model-fitting is compared to that of the second-order LPPL model in Fig. 5.4. This second-order model fit is recreated from the parameters given in Johansen & Sornette (1999b), namely,

$$t_c = 1989.97, \quad \beta = 0.41, \quad \omega = 4.8, \quad \Delta_t = 9.5, \quad \Delta_\omega = 4.9. \tag{5.39}$$

One can see that the logistic model fits the 5.5 years of data very well, as does the second-order model. Fig. 5.4 shows how the observations develop compared to the predictive progression of the fitted models out until 2002.5. Although neither of the models provide a meaningful fit to the future, non-fitted data, one can see that as the second-order model saturates it cannot follow the steepening downward trend of the time series. Conversely, as $t \to t_c - \Delta_t$, the logistic model begins to oscillate more rapidly and tends to the unbounded downside. As will be seen, this seems to more accurately match the actual development of this anti-bubble.

Second, the 9 years of observed data from the Nikkei 225 index from 1990 is considered, and the results of the logistic model are compared with that of the third-order LPPL model.

Following the same method as above, consecutive pairs of troughs prior to

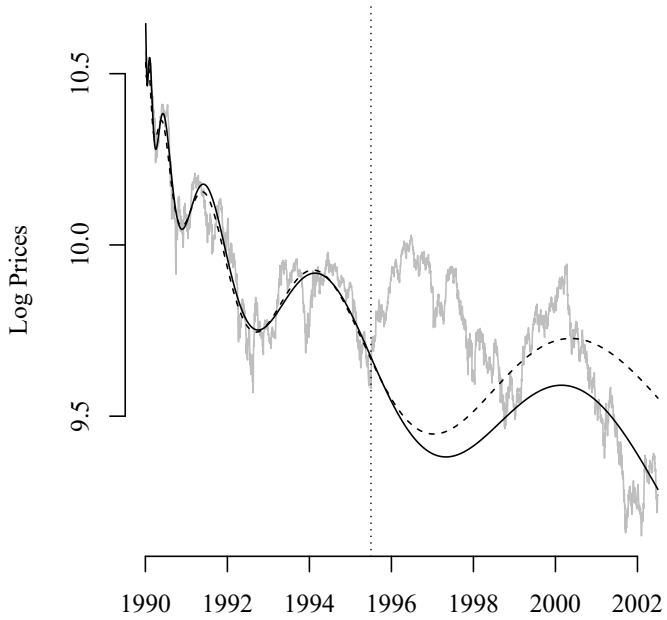


Figure 5.4: Second-order model using Johansen & Sornette (1999b) parameter fit (dashed line) and logistic models (solid line) using posterior mean of each parameter from MCMC simulation as in (5.38). The data set used is between 1990–1995.5, the end of this period being marked by a vertical dotted line. As can be seen the second-order and logistic models fits are similar for this period. However, when the time series is extended out to 2002.5, at first sight the logistic model seems to have more accurate long-term predictive properties.

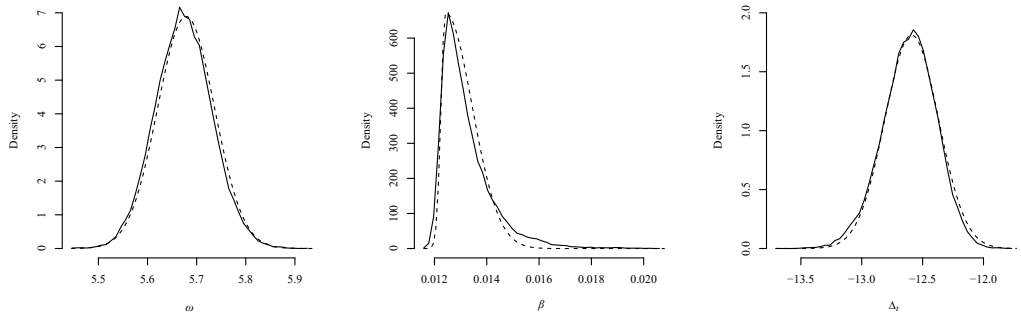


Figure 5.5: Density plots of posterior distributions for logistic model fitted between 1990–1999 with approximate normal distributions (dotted lines) for the ω and Δ_t parameters. The parameter β is approximately skew-normal distributed.

1999 are taken from Table 5.1 that give negative values of Δ_t. The set of pairs are shown in Table 5.3 and give $\bar{\Delta}_t = -10.98$ and $s^2_{\Delta_t} = 3.87$, and $\bar{\omega}_t = 5.02$ and $s^2_{\omega_t} = 0.69$. Again, using the prior distributions in (5.37), the MCMC simulation produces approximately normal posterior distributions for two of the parameters (ω and Δ_t), and a skew-normal distribution for the third parameter β, as follows

$$\beta \sim SN(0.013, 0.001, 2.5), \quad \omega \sim N(5.68, 0.06), \quad \Delta_t \sim N(-12.59, 0.22)\,. \tag{5.40}$$

Again, the results of this model fitting are compared to that of a LPPL model, in this case, the third-order model. This is shown in Fig. 5.6. This third-order model fit is again recreated from the parameters given in Johansen & Sornette (1999b). For this 9 year plot, the authors retain the parameters t_c, β and ω from the previous 5.5 year second-order fit and vary only Δ_t, Δ_ω, $\Delta_t^{'}$ and $\Delta_\omega^{'}$. The curve is fitted with the following values

$$\Delta_t = 4.34\,, \quad \Delta_\omega = -3.10\,, \quad \Delta_t^{'} = 7.83\,, \quad \Delta_\omega^{'} = 23.4 \tag{5.41}$$

Fig. 5.6 shows the third-order model more accurately fitting the amplitudes of the observed data that the logistic model. However, the logistic model picks out the periodicity of the observed data equally as well as the third-order model, and it manages to do this with three fewer nonlinear parameters. With both the

t_1	t_2	t_3	t_4	ω	Δ_t
1990.20	1990.70	1992.60	1995.50	5.49	−11.35
1990.20	1990.70	1995.50	1998.80	5.51	−12.28
1990.70	1992.60	1992.60	1995.50	3.98	−7.76
1990.70	1992.60	1995.50	1998.80	4.27	−10.75
1992.60	1995.50	1995.50	1998.80	5.84	−12.78

Table 5.3: Implied values for ω and Δ_t from pairs of consecutive troughs suitable for data between 1990–1999 excluding positive values of Δ_t.

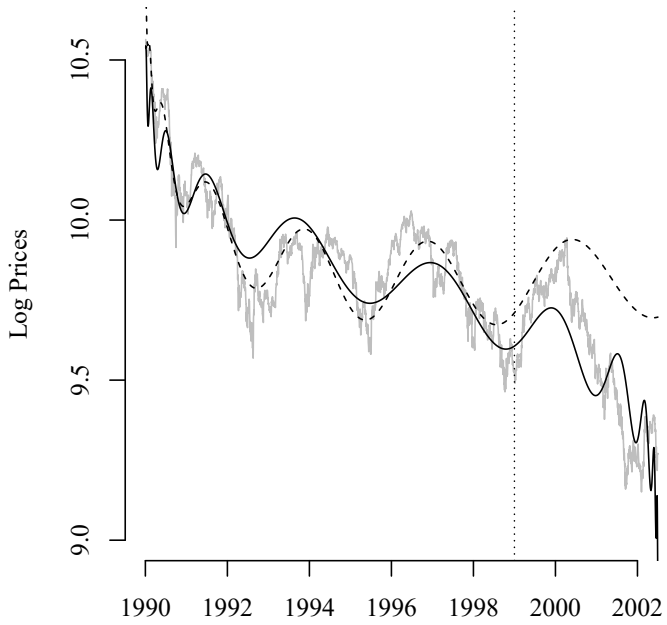


Figure 5.6: Third-order model using Johansen & Sornette (1999b) parameter fit (dashed line) and logistic models (solid line) using posterior mean of each parameter from MCMC simulation. The data set used is between 1990–1999, the end of this period being marked by a vertical dotted line. Note the remarkable predictive quality of the logistic model when the data series is extended to future, non-fitted data between 1999–2002.5. Having a singularity at $t_c + |\Delta_t|$ (for $\Delta_t < 0$), the logistic model does not apply past the end of the anti-bubble.

second- and third-order models, the fits to the data are generally good because as t moves away from t_c their angular log-frequencies are constantly changing. Conversely, the logistic model is able to track the periodicity by holding is log-frequency constant but assuming a second singularity at a distance Δ_t from t_c.

The striking feature of the logistic model is how the progression of the model closely follows the development of the non-fitted future data over the subsequent 3.5 years. By this time the third-order model is beginning to saturate and offers no predictive value past the beginning of 2000. However, the logistic model's increasingly rapid oscillations and steep descent describe the path of this index very well to the naked eye. Of course, this is an out of sample fit rather than a prediction, and it should be noted that in May 1999 the third-order LPPL model produced a very accurate ex-ante prediction of the trend reversal in the Nikkei index over the subsequent year (Johansen et al. 2000).

Furthermore the 95% confidence limit for the distribution of Δ_t is between -12.18 and -13.05 implying an end to the anti-bubble between 2002.18 and 2002.05. As it happened, the anti-bubble came to an end in the the first half of 2003, around 2003.3.

Finally, Fig. 5.7 shows a fit of the logistic model using the method employed for the previous two examples over the whole time period 1990–2003. Given the saturating features of the second- and third-order LPPL equations, one would not expect these models to perform well, but, intriguingly, the time series seems to suit the logistic model extraordinarily well.

5.5 Summary and interpretation of results

This chapter has shown how the suite of log-periodic power-law models for asset bubbles and anti-bubbles can be extended to include models derived from solutions to the Schröder equation which satisfy the original renormalisation group equations. In particular, models have been derived from the logistic differential equation, which have been called the logistic model and modified logistic model. No claim has been made that these models describe reality better (or worse)

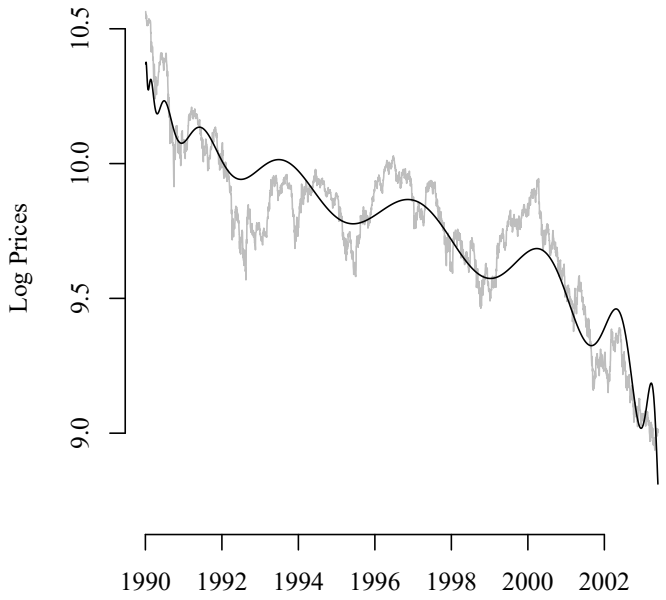


Figure 5.7: The logistic model is fitted from 1990 to the beginning of 2003. Although the amplitude of the oscillations is not accurately described by the model the periodicity seems to be very good. Additionally, the found value of Δ_t implies an end to the anti-bubble at 2003.66. As stated previously the Nikkei 225 began its upswing shortly before this around 2003.3

than the established models, but they appear to be important additions to the modelling toolbox.

Indeed, on examining the S&P500 index bubble between 1980 and the crash of October 1987, one can observe that the logistic model may have some advantages in terms of flexibility over the first-order LPPL model in situations where there is a required value for the critical time, and where, on first sight, the observed data appears to have oscillations with varying log-frequencies. Additionally, the logistic model has the advantage of fewer parameters over the second- and third-order extensions.

In the case of the Nikkei 225 index anti-bubble between 1990 and 2003, the addition of a second singularity in the logistic model derived by taking negative values of Δ_t provides a model with features that are remarkably predictive over long time-periods for this particular data set.

Generally, the use of solutions of the renormalisation group formalism corresponding to a nonlinear $\phi(x)$ is appealing, as is the implied beginning to the bubble (and end to the anti-bubble) given by $\Delta_t < 0$. However, this chapter has not gone far in investigating whether or not these ideas have any concrete advantages over any other method of modelling bubble or anti-bubble regimes. As a development of this work, one could take a more rigorous approach to the determination of the logistic model's parameters over a wide range of faster-than-exponential growth/decline regimes, and discover whether or not there is any value in extending these models with further perturbations as has been done in the derivation of the modified logistic model. A key test of any new models will be whether they can successfully make *ex ante* predictions of the end of an asset bubble and/or anti-bubble.

As a last note on the logistic model, another interesting feature of the model fit to the S&P500 index between 1980 and the October 1987 crash is that the fitted value of $\Delta_t = -15.7$ implies a singularity $I(t) \to -\infty$ when $t \to 1972.1$. Clearly this feature is not seen in the observed data, and by plotting the model further back in the time series, the logistic model provides no particular insight into the pre-1980s development of the bubble, as we have seen in the predictive

aspects of the logistic model when fitted to the Nikkei 225 index anti-bubble between 1990–1999. This suggests that anti-bubbles cannot simply be regarded as bubbles in reverse time, but have their own distinctive characteristics and dynamics.

The question that remains is how can these models be used from a practical perspective and is there a method that relies less on the large parameter space to detect DSI? The following chapter changes direction to investigate methods to fit change-point models based on simple power laws, to both the pre- and post-crash time-series to understand better the onset of asset price bubbles.

Chapter 6

Detecting discrete scale invariance

As can be seen from the previous section, fitting LPPL models to observed data is certainly not straightforward, and it is a matter of debate whether doing so leads one closer to demonstrating that the observed data is really displaying any degree of discrete scale invariance. With this in mind, the question can be asked whether there are any other methods of detecting DSI that are less reliant on models with such sloppiness as described in Bree et al. (2013)?

The first method is based on the proposal that the price behaviour in financial markets could be generally in one of three states, i) constant rate of return or decline, ii) super-exponential return, a "bubble", or iii) super exponential decline, which is either a "crash" or an "anti-bubble", a crash in this context being a draw-down exceeding a defined amount over a defined period. The model describing the superexponential return or decline is expressed as a simple power law in the log-prices of the asset, and the angular log-frequency is ignored. It is supposed that there are points where returns switch between regions that are characterised by one or other of these regimes (the "change points"), and that a constant-rate-of-return regime can only switch to a superexponential growth regime, and such a superexponential growth regime must be followed by period of crash or anti-bubble. Furthermore, it is assumed this period of decline is followed by a period of constant returns, and a piecewise model of this behaviour is derived.

Secondly, and most importantly, a revised methodology to identify change-points in financial markets where the governing regime shifts from a constant rate-of-return, i.e. normal growth, to superexponential growth described by, in this case, a power-law hazard rate in a stochastic model of the log-price returns of the observed data rather than the log prices themselves. The latter regime corresponding to financial bubbles driven by herding behaviour of market participants. Assuming that the time series of log-price returns of a financial index can be modelled by arithmetic Brownian motion, with an additional jump process with power-law hazard function to approximate the superexponential growth, a threshold value of the hazard-function control parameter is derived, allowing a decision to be made regarding in which regime the market is more likely to be at any given time. An analysis of the S&P500 index over the last 60 years provides evidence that the methodology has merit in identifying when a period of herding behaviour begins, and, perhaps more importantly, when it ends.

6.1 Piecewise regime change model

Returning to the proposal that a piecewise model may be derived of segments of unknown length comprised of combinations of regimes of constant rate of return or decline, superexponential return, superexponential decline, one could define a model that switches between regimes at change points with time $t = c_1$, where the market moves from a region of constant return to that described by the superexponential return of a power law, and $t = c_2$, where the bubble collapses and a crash typified by a period of power law ensues, and then, finally, the constant-rate-of-return regime is restored at $t = c_3$. This being the case, the whole model of the asset log price, $f(t)$, may be expressed as the piecewise equation

$$f(t) = \begin{cases} f_1(t) & t \leq c_1 \\ f_2(t) & c_1 \leq t \leq c_2 \\ f_3(t) & c_2 \leq t \leq c_3 \\ f_4(t) & t \geq c_3 \end{cases} \tag{6.1}$$

Assuming that stock price movements are intra-day continuous, the equations $f_n(t)$ can be written as

$$f_1(t) = a_1 + b_1 t \tag{6.2}$$

$$f_2(t) = f_1(c_1) + b_2\left((t_c - t)^{\beta_1} - (t_c - c_1)^{\beta_1}\right) \tag{6.3}$$

$$f_3(t) = f_2(c_2) + b_3\left((t - t'_c)^{\beta_2} - (c_2 - t'_c)^{\beta_2}\right) \tag{6.4}$$

$$f_4(t) = f_3(c_3) + b_4(t - c_3) \tag{6.5}$$

where t_c is the critical time and β_1 is the critical exponent of the bubble regime, and similarly t'_c is the critical time of the anti-bubble regime, and β_2 is its critical exponent.

This derives a model with a twelve dimensional parameter space that can be used to fit a time-series of observed log-price data. However, some restrictions must be placed on the parameters so that they make sense from the perspective of this data. Firstly, it is clear that the change point c_1 must precede c_2, which in turn must precede c_3. Secondly, if the change point c_2 is the point at which a crash actually occurs, it is known that the critical point, t_c, must occur subsequent to this point. Similarly, for the crash or anti-bubble starting at the change point c_2, it is known that the associated critical point, t'_c, must occur prior to this point. The critical exponents must be such that at the critical points in the time series are finite but singular in their rate of change. Therefore, β_1 and β_2 must lie between 0 and 1. Finally, since one is looking for particular occurrences of market bubble formation between c_1 and c_2 we can expect that $f_1(c_1) \leq f_2(c_2)$. Clearly then, $b_2 \leq 0$. Similarly, it is expected that after the crash between the change points c_2 and c_3, it is necessary that $f_2(c_2) \geq f_3(c_3)$. This implies $b_3 \leq 0$.

The question to be answered is whether by fitting this model to a large data set of financial data one can uncover evidence of this model being any more "valid" in known times of bubble and crash than in more stable times. To answer this question, a MCMC simulation was built of the model to examine how the convergence of the simulation performs when presented with a very large data set of time-series of varying lengths, testing whether periods exist where the model

showed strong convergence coinciding with known market events, in comparison to periods during which the model does not converge.

To create such a model one must decide the most appropriate prior distributions from which to sample the model parameters. Unfortunately, there is very little prior information to hand regarding the value of the parameters. Firstly, there are five points of interest in the time series, namely the time of the first observation, c_0, followed by the three change points c_1, c_2, and c_3 described above, and the time of the last observation, c_4. c_0 and c_4 are fixed constants but the remaining points are chosen from a uniform distribution, $\mathrm{U}(a, b)$ where any point between a and b is equally likely, such that

$$c_1 \sim U(c_0, c_4), \quad c_2 \sim U(c_1, c_4), \quad c_3 \sim \mathrm{U}(c_2, c_4) \tag{6.6}$$

Furthermore, the two critical exponents, β_1 and β_2 must be chosen such that

$$\beta_1 \sim U(0, 1) \quad \beta_2 \sim U(0, 1) \tag{6.7}$$

Furthermore, an assumption is made regarding the critical points, t_c and t'_c in that they are somewhere close (respectively subsequent or preceding), and chosen to be within 0.1 years to the change point c_2, so that

$$t_c \sim U(c_2, c_2 + 0.1) \quad t'_c \sim U(c_2 - 0.1, c_2) \tag{6.8}$$

These last two assumptions are somewhat arbitrary, and on reflection, one could argue that it would be a reasonable simplification to make the direct assumption that the two critical times should be fixed at the point c_2.

Finally, the prior distributions for the parameters a_1 and b_n must be specified, where $n = 1, 2, 3, 4$. At first sight, this seems to be rather difficult as we have no knowledge of where these parameters should be set other than the rather general restrictions set out previously. However, sensible prior distributions of the model log-prices at each of the points, c_n, can be arrived at. By taking the sample mean of the log-prices, $\bar{y}$ and sample variance, σ_y^2 we take the prior distribution of the

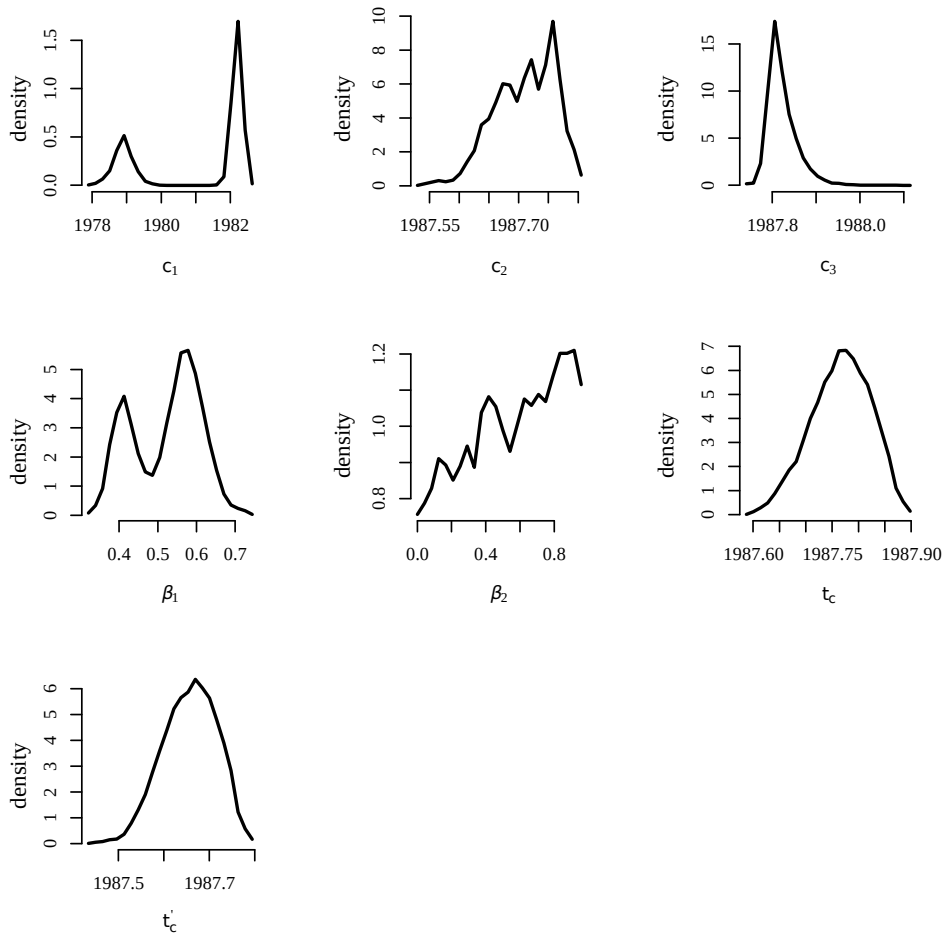


Figure 6.1: Density plots of each change point model parameter applied to the S&P500 index during the late 1970s and 1980s. One can see convergence issues in the plots of c_1 and β_1, but the other parameters have seemed to converge reasonably. The most likely value of the change point parameters c_2 and c_3 fit the observed data very well. The most likely value of the change point c_1 is around 1982, but the distribution has another peak around 1979. A visual inspection of the time-series would suggest that the log-periodic oscillations began around 1980.

fitted value at time c_0 to be normally distributed as $f(c_0) \sim N(\bar{y}, {\sigma_y}^2)$. The same approach can be taken for $f(c_1)$ and $f(c_4)$. However, the distribution of the fitted value $f(c_2)$ is taken from the same normal distribution but truncated to take samples greater than the observed value at c_1, i.e. the model is looking for accelerated returns in the bubble region, and as such the fitted value at c_2 must be higher than that at c_1. Similarly, since for the model to be valid there should be a collapse in prices after time c_2, the distribution for $f(c_2)$ is truncated to take samples only below that of the observed value at time c_2. Having derived prior distributions for model values at the points c_n, the parameters a_1 and b_n can be solved as

$$\begin{aligned} a_1 &= \frac{c_0 f(c_1) - c_1 f(c_0)}{c_0 - c_1} \\ b_1 &= \frac{f(c_0) - f(c_1)}{c_0 - c_1} \\ b_2 &= \frac{f(c_1) - f(c_2)}{(t_c - c_1)^{\beta_1} - (t_c - c_2)^{\beta_1}} \\ b_3 &= \frac{f(c_2) - f(c_3)}{(c_2 - t_c')^{\beta_2} - (c_3 - t_c')^{\beta_2}} \\ b_4 &= \frac{f(c_3) - f(c_4)}{c_3 - c_4} \end{aligned} \tag{6.9}$$

The likelihood of the model fitting the data given the parameters is now assumed to be drawn from a normal distribution, the mean of which is given by equation (6.1).

A first impression of this method is that some interesting results may be produced when looking at specific market events. For example, Fig. 6.1 shows density plots for each parameter when the change point model is fitted to the S&P500 index between 1975–1993. One can see from the diagram that there are some convergence issues which need to be rectified, i.e. the multiple peaks are indicative that the MCMC chains have not converged. Interestingly, the change point model seems to be undecided whether the first change point transition from a constant growth to power-law acceleration, c_1, occurs around 1979 or 1982. It would seem there is much more certainty regarding the change points c_2 and c_3, the location of the crash and the transition back to constant growth respectively.

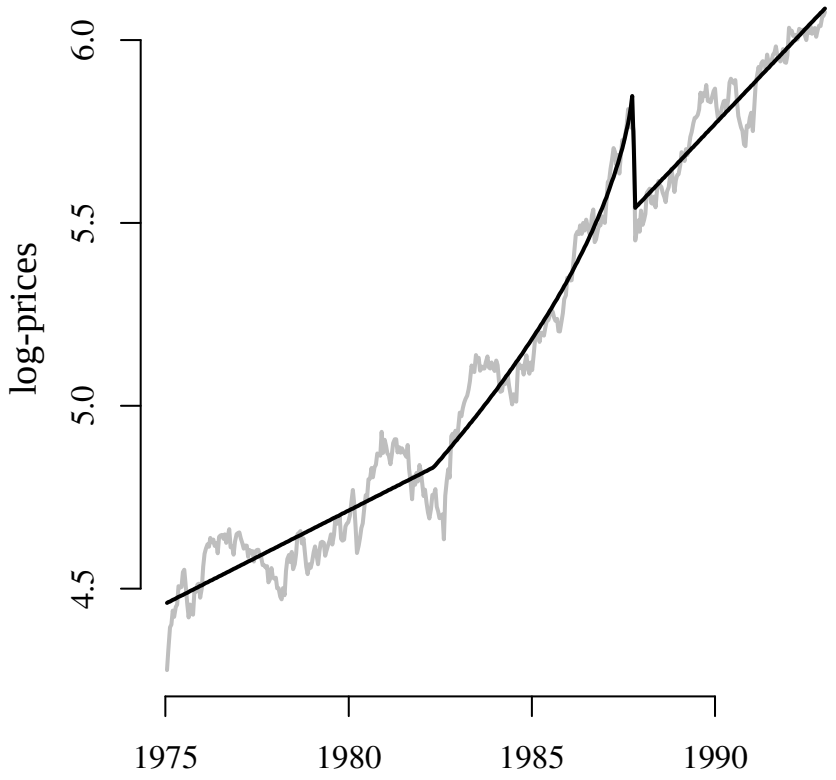


Figure 6.2: Change point model fitted to S&P500 index between 1975–1992. The black line shows the fitted values of the change point model as compared to the actual index in grey. The fit of this time series is quite striking, and the start of the exponential growth period appears to be around 1982.

Fig. 6.2 is a plot of the change point model using the most likely value of each parameter in comparison to the index. As one can see, the model fits the observed data very well, particularly in respect of describing the power law acceleration prior to the crash.

As a counterpoint to this successful outcome, one can see a different story when the S&P500 index between 1995–1998 is examined. The change point model is fitted in the same manner as before, but during this period there is no crash event. As can be seen from Fig. 6.3, the MCMC simulations have not converged well for the change points c_2 and c_3, although it could be said that there is good convergence for change point c_1.

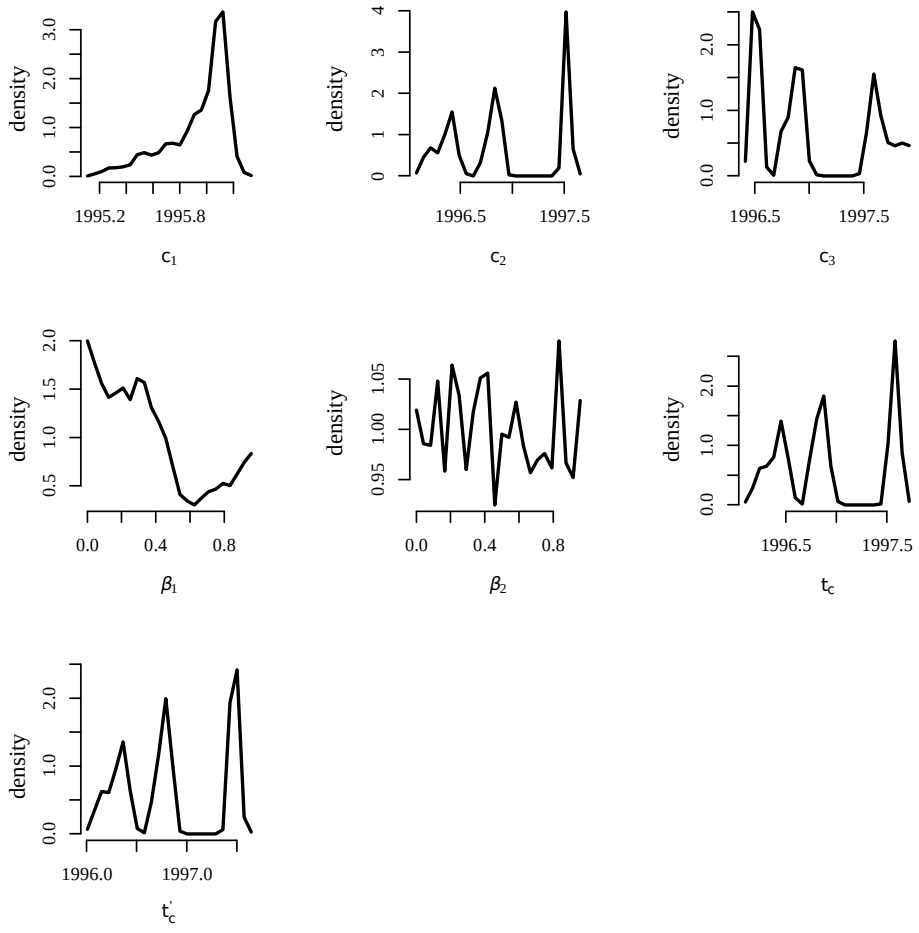


Figure 6.3: Density plots of the change point model parameters for the period between 1995–1998 of the S&P500 index. The MCMC simulation with 7,500 iterations has produced much less stable parameter estimation for a relatively benign market.

Furthermore, when the most likely values of the parameters are used to compare the model with the actual index, as in Fig. 6.4, one can see that the fitted values do not describe the observed data well.

This result is in line with expectations, i.e. during periods in which the change point model is valid, there should be good convergence in the main change point model parameters. However, in times where the underlying mechanics driving the market do not follow a power-law, simulating these periods with the change point model should yield poor convergence and fitting results.

It would seem that in isolated instances, this method goes some way to identifying those points where a bubble begins, how it grows, and when and how it ends. The method is worthy of more research, but, in practice it can only be be used in very limited circumstances and only ever from an *ex post* perspective. Additionally, an assumption is made that a bubble is always followed by a crash or period of anti-bubble. As will be seen in the next section, when one looks at the JLS model of rational expectations, it is clear that from a probability perspective, it is certainly not necessary that a bubble always ends in a crash. Therefore, although this analysis is perhaps interesting in of itself, there are other approaches that can be constructed that are more agnostic in terms of the structural prior assumptions and less reliant on a large parameter space. Just such a method is described in the following section and lays down the theory for an *ex ante* predictive model that may be used in real time.

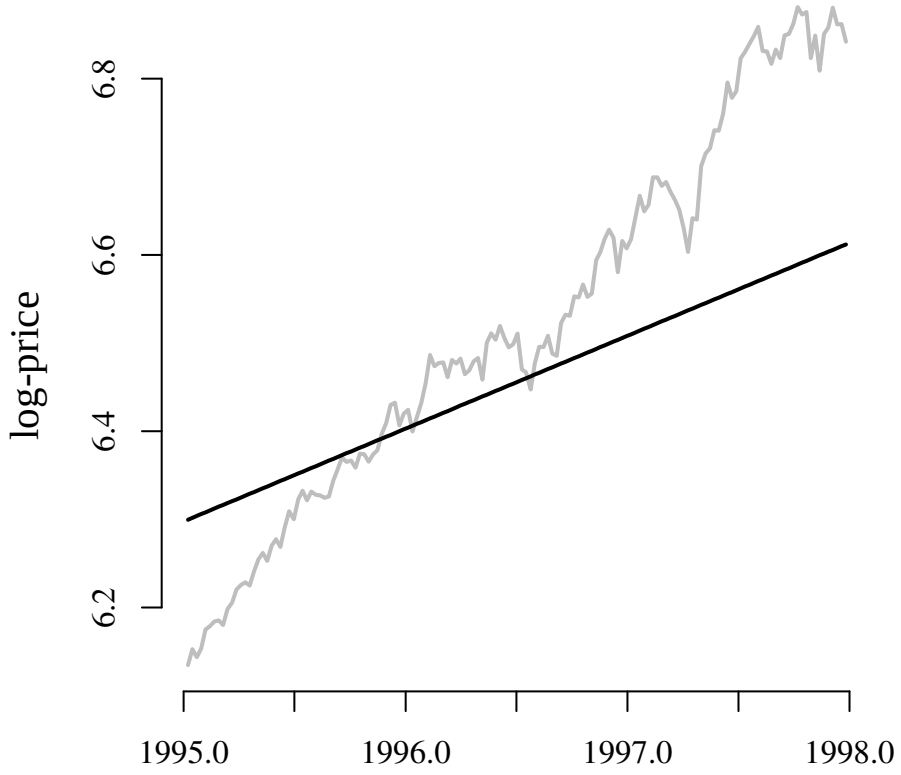


Figure 6.4: Change point model fitted to S&P500 index between 1995–1998. The black line shows the fitted values of the change point model as compared to the actual index in grey. During this period of the model fit is very poor, possibly indicating there are no periods of regime change during this period.

6.2 Change-point model for asset price bubbles with power-law hazard function.

When studying asset-price bubbles and anti-bubbles, a holy grail of theoreticians and practitioners alike is to predict their start and end, preferably as early as possible, so as (according to circumstances and inclination) to take maximum advantage of any opportunities that might arise or to mitigate any negative consequences or even to intervene in the market to prevent or curtail such a bubble/anti-bubble. As a result there has been a large amount of literature devoted to the detection of bubbles/anti-bubbles, with several competing approaches (see Gürkaynak 2008 and Jarrow 2016).

The valuation of stock markets, and of financial assets more generally, has been the subject of much debate. For some who favour the efficient markets hypothesis (in its various guises, including the random-walk model), bubbles fail to exist, full market knowledge having been already factored into the asset price, either immediately or sufficiently quickly to prevent any effective use of temporary market departures from the fundamentals.

In this section, as in the preceding sections, it is assumed that certain periods of accelerated superexponential growth may be viewed as financial bubbles driven by herding behaviour of market participants. Indeed, for many finance researchers and industry practitioners, the existence of asset bubbles and anti-bubbles is an established fact, a commonplace occurrence, and, indeed, many economists and practitioners question the strict application of the efficient markets hypothesis. For example, in Robertson & Wright (1998), the authors report that long-term stock returns appear to be "much less uncertain than a random-walk model would imply," and that research suggests that there is a "weak tendency for stationary valuation indicators to predict future stock prices" so that "long-run returns can become markedly more predictable".

Furthermore, there is a wide variety of methods used to model change-points in time series, depending on the application and research community. For example, for the econometrics community, the detection of change-points between

bubbles/anti-bubbles may proceed by fitting standard time-series models such as ARIMA, (G)ARCH (Mills & Markellos 2008) and determining time points at which the nature of the models change, often using Dickey-Fuller-type tests to detect unit roots in the underlying statistical models. Several review papers discuss this approach, see, for example, Taipalus (2012), Arshanapalli & Nelson (2016), Harvey et al. (2015), Astill et al. (2017) and references therein.

There is a large body of literature in the statistics community on change-point methods, including MCMC methods, e.g. Adams & MacKay (2007), Benson & Friel (2018), Heard & Turcotte (2017). See Aminikhanghahi & Cook (2017) for a review from the machine-learning community. The so-called Pruned Exact Linear Time (PELT) algorithm (Killick et al. 2012), a development of the method of Jackson et al. (2005), which uses both dynamic programming and pruning to produce an algorithm that is linear in the data size, is included in the R package (Killick & Eckley 2013), along with the Segment Neighbourhood algorithm, and the classical Binary Segmentation algorithm.

The approach to change-point identification taken in this section is somewhat different from these methods. First, a normal and log-normal bi-variate distribution of growth rate and volatility is obtained empirically by fitting data from the times series over random time periods. Second, starting from an assumption that the asset grows exponentially following a geometric Brownian motion process, with superexponential growth modelled by an additional power-law hazard term (activated when a parameter $\nu \neq 0$ following Cheah & Fry (2015)), the model is simulated using the growth rates and volatility selected randomly from the previously determined bi-variate distribution. To test whether the data fits the superexponential growth model, a most-likely-value estimate (**MLV**) is obtained from the distribution of ν calculated from a maximum-likelihood estimate based on random simulation of the other model parameters. In the case studied, these MLV estimates are found empirically to be well fitted by generalised/skew logistic distributions (for both normal and superexponential growth). The two resulting distributions are close but distinct and provide signatures for the two types of growth.

These distributions are then used as proxies for the probabilities of superexponential and normal growth themselves, and a threshold is determined which is considered as the boundary between these two regimes. Using these signature distributions, it is possible, using past and present prices, to decide, for each trading day, which is more likely to be the governing regime: superexponential growth or normal growth.

This section is divided as follows: in §6.2.1 a model of returns during a financial bubble as an arithmetic Brownian motion stochastic differential equation modulated by a jump process governed by a power-law hazard function is presented, and the control parameter ν is described, closely following Cheah & Fry (2015). In §6.2.2 these defined models of normal and superexponential growth are used to simulate time series, and find the threshold value of the control parameter, ν, under each regime. In §6.2.3, this threshold value is applied to the last 60 years of data from the S&P500 index and the change-points between bubble and normal market regimes are calculated. This section concludes with some final remarks in §6.2.4.

6.2.1 Modelling financial markets with power-law hazard rate

Asset price bubbles as a stochastic process

In deriving a model of periods of superexponential returns, it makes sense to initially define a model for markets operating in "normal" conditions. The first assumption is that returns on an asset-price index are log-normally distributed, and, as such, the price S_t can be modelled by the geometric stochastic differential equation (Hull 2017)

$$dS_t = \left(\mu_t + \frac{\sigma_t^2}{2}\right) S_t \, dt + \sigma_t S_t \, dW_t \,, \tag{6.10}$$

where $\mu_t + \sigma_t^2/2$ is the drift coefficient, σ_t is the diffusion coefficient, and W_t is a standard Wiener process with $E[W_t] = 0$ and $E[W_t^2] = t$. Writing $X_t = \log S_t$ and then, by applying Itô's formula (being in essence a chain rule for stochastic

functions)

$$dX_t = \left(\frac{\partial X_t}{\partial S_t} \left(\mu_t + \frac{\sigma_t^2}{2} \right) S_t + \frac{\partial X_t}{\partial t} + \frac{\partial^2 X_t}{\partial S_t^2} \frac{\sigma_t^2}{2} S_t^2 \right) dt + \frac{\partial X_t}{\partial S_t} \sigma_t S_t dW_t \tag{6.11}$$

Since,

$$\frac{\partial X_t}{\partial S_t} = \frac{1}{S_t}, \quad \frac{\partial^2 X_t}{\partial S_t^2} = -\frac{1}{S_t^2}, \quad \frac{\partial X_t}{\partial t} = 0 \tag{6.12}$$

it can be seen that

$$\begin{aligned} dX_t &= \left(\frac{1}{S_t} \left(\mu_t + \frac{\sigma_t^2}{2} \right) S_t - \frac{1}{S_t^2} \frac{\sigma_t^2}{2} S_t^2 \right) dt - \frac{1}{S_t} \sigma_t S_t dW_t \\ &= \left(\left(\mu_t + \frac{\sigma_t^2}{2} \right) - \frac{\sigma_t^2}{2} \right) dt - \sigma_t dW_t \end{aligned} \tag{6.13}$$

so that one obtains the arithmetic Brownian motion stochastic differential equation

$$dX_t = \mu_t \, dt + \sigma_t \, dW_t \, . \tag{6.14}$$

Let the event Y be the occurrence of a market crash at time $t = t_Y$. Following Johansen et al. (2000) and Cheah & Fry (2015), a conjecture is made that in the period leading up to a crash, the asset price follows the same process as in (6.10), but suffers a deterioration in the price by a factor κ, where $0 < \kappa < 1$, at the time of a crash, t_Y. Therefore, the asset price in this regime, $\tilde{S}$, follows the process

$$\tilde{S}_t = S_t (1 - \kappa)^{j_t} \tag{6.15}$$

where j_t is the jump process

$$j_t = \begin{cases} 0 & t < t_Y \, , \\ 1 & t \geq t_Y \, . \end{cases} \tag{6.16}$$

Now, writing $\tilde{X}_t = \log(\tilde{S}_t) = \log S_t + j_t \log(1 - \kappa)$ so that, applying Itô's formula again with

$$\frac{\partial X_t}{\partial S_t} = \frac{1}{S_t}, \quad \frac{\partial^2 X_t}{\partial S_t^2} = -\frac{1}{S_t^2}, \quad \frac{\partial X_t}{\partial t} = \frac{dj_t}{dt} \log(1 - \kappa) \tag{6.17}$$

gives a modified stochastic differential equation

$$d\tilde{X}_t = \mu_t\, dt + \sigma_t dW_t + \log(1-\kappa)\, dj_t\,. \tag{6.18}$$

The hazard rate

Under the condition that event Y has not occurred up to time $t < t_Y$, the jump process $j_t = 0$ and its expected value at time t, $\mathrm{E}[j_t] = 0$. Therefore, for the infinitesimally small time period δt,

$$\mathrm{E}[j_{t+\delta t} - j_t] = \mathrm{E}[j_{t+\delta t}]\,. \tag{6.19}$$

The expected value of $j_{t+\delta t}$ can be interpreted as the probability there is a crash event, Y, in the interval $[t, t+\delta t]$ conditioned on there not having been such a event up until this point. If the hazard rate, h_t, is defined as the probability per unit time that a crash event, Y, occurs in the next moment, δt, conditioned on it not having already happened, then

$$\mathrm{E}[j_{t+\delta t}] = h_t \delta t \tag{6.20}$$

Similarly, the variance of j_t conditioned on the event Y not having occurred up to time t is given by $\mathrm{Var}[j_t] = 0$. Therefore, since it is clear that $j_t = j_t^2$,

$$\mathrm{Var}[j_{t+\delta t}] = h_t \delta t - h_t^2 \delta t^2\,. \tag{6.21}$$

Expectation and variance of returns in the hazard rate model

The return of an asset in the time interval $[t, t+\delta t]$ is given by $d\tilde{X}_t = \tilde{X}_{t+\delta t} - \tilde{X}_t$. Since $\mathrm{E}[dW_t] = 0$, it follows from (6.18) that, in a bubble regime, the expected value of the asset return in $[t, t+\delta t]$ is

$$\mathrm{E}[\tilde{X}_{t+\delta t} - \tilde{X}_t] = \mathrm{E}[\mu_t \delta t] + \log(1-\kappa)\mathrm{E}[j_{t+\delta t} - j_t]\,. \tag{6.22}$$

Therefore, from (6.20) it can be seen that

$$\mathrm{E}[\tilde{X}_{t+\delta t} - \tilde{X}_t] = \mu_t \delta t + \log(1-\kappa) h_t \delta t \,. \tag{6.23}$$

Now, the expected rate of return, $\mathrm{E}[\tilde{X}_{t+\delta t} - \tilde{X}_t]$, can be interpreted as the mean return over all periods of δt across the whole time series. Following Cheah & Fry (2015), this rate of return is assumed to be a fixed value per unit time, μ, so that

$$\mu_t = \mu - \log(1-\kappa) h_t \,. \tag{6.24}$$

Since $\log(1-\kappa) < 0$, Cheah & Fry (2015) interprets this to suggest that, as the risk of a crash occurring increases, the return required for traders to stay in the market must increase to compensate for the increased risk.

By a similar argument, assuming a fixed volatility σ across the whole time series, Cheah & Fry (2015) gives the relationship between the hazard function and the volatility as

$$\sigma_t^2 = \sigma^2 - (\log(1-\kappa))^2 h_t \tag{6.25}$$

implying, as the authors remark, the rather counter-intuitive suggestion that as the risk of a crash increases, the volatility of the asset price actually reduces, perhaps an indication of overconfidence in the market as bubbles mature.

The theory in Cheah & Fry (2015) is valid for finite hazard rate. However, when, as in the next section, the hazard rate is modelled as a power law with $h_t \to \infty$ as t approaches a critical time t_c, it is clear that equation (6.25) is problematic close to t_c because then $\sigma_t^2 < 0$. Note that in Johansen et al. (2000) it is assumed that $\sigma_t = 0$ so that the stochasticity enters through the jump process, while in Cheah & Fry (2015) the full model (6.18) is presented. This thesis takes a middle position, modelling $\mu(t)$ through (6.24), but making the ansatz $\sigma_t = \sigma$, a non-zero constant. While it is clear future theoretical developments might incorporate time-varying σ_t, the justification for the ansatz is twofold: *ex ante* because the focus of this thesis is on the modelling of the growth rate μ_t, and *ex post* because, as will been seen in §6.2.2, the ansatz allows the construction

of a statistic that appears to discriminate 'normal' exponential growth from the superexponential growth that is characteristic of asset bubbles.

The hazard rate as a power-law

Asset bubbles (and anti-bubbles) result from imitative behaviour or herding in the market. Restating the theory discussed in earlier chapters, the likelihood of this imitation is a function of the general interconnectedness of the market participants, such that if δ is the number of connections of a typical trader, one might suppose that the hazard rate satisfies $dh_t/dt = ch_t^\delta$ with $\delta > 1$ and c constants. Solving the differential equation with this condition gives

$$h_t = B(t_c - t)^\beta, \quad \beta = -\frac{1}{\delta - 1} < 0 \tag{6.26}$$

where t_c is a constant for which $h_t \to \infty$ as t approaches t_c, the critical time. It is important to keep in mind that the critical time, t_c is considered as not being the actual time of the crash, but rather the time when the crash is most likely to happen (Johansen et al. 2000). Furthermore, one takes $\delta > 2$ so that $-1 < \beta < 0$, which is required for the asset to have a finite value but infinite derivative at $t = t_c$.

If t_Y is the time of a proposed crash, then the probability that the crash occurs in the time interval $(t_1, t+\delta t)$ is given by $P(t_1 \leq t_Y \leq t+\delta t)$. If it is assumed that the probability of a crash not occurring in either of the non overlapping intervals (t_1, t) and $(t, t+\delta t)$ then clearly

$$P(t_1 \leq t_Y \leq t + \delta t) = P(t_1 \leq t_Y \leq t) \times P(t < t_Y \leq t + \delta t)\,. \tag{6.27}$$

Since, $P(t < t_Y \leq t + \delta t) = 1 - h(t)\delta t$ it can be seen that

$$P(t_1 \leq t_Y \leq t + \delta t) = P(t_1 \leq t_Y \leq t)(1 - h(t)\delta t) \tag{6.28}$$

which gives the relationship

$$-h(t)\delta t = \frac{P(t_1 \leq t_Y \leq t+\delta t) - P(t_1 \leq t_Y \leq t)}{P(t_1 \leq t_Y \leq t)} \tag{6.29}$$

so that, when allowing $G(Y_t) = P(t_1 \leq t_Y \leq t)$ in the limit $\delta t \to 0$, the above expression becomes

$$-h(t)dt = \frac{dG(Y_t)}{G(Y_t)}\,. \tag{6.30}$$

Then it is straightforward to see that the probability distribution function, $G(Y_t)$, of a crash not occurring in the interval (t_1, t), where t_1 is the initial time, is given by

$$G(Y_t) = \exp\left[-\int_{t_1}^{t} h_\tau \, d\tau\right] \tag{6.31}$$

Therefore, when $t < t_c$,

$$G(Y_t) = \exp\left[\frac{B}{1+\beta}\left((t_c - t)^{\beta+1} - (t_c - t_1)^{\beta+1}\right)\right] \tag{6.32}$$

and β should be restricted such that it lies within the interval $[-1, 0)$ to prevent $G(Y_{t_c})$ from being singular. Therefore, and it seems reasonable to suggest, there is a finite probability that the bubble ends without a crash at all. Furthermore, it follows that since it is required that $G(Y_t)$ must lie in the interval $[0, 1]$, there is the additional restriction $B \geq 0$. If one writes $\nu = B\log(1-\kappa)$, then $\nu \leq 0$ for $0 < \kappa < 1$.

Now expressions of the statistical parameters for a market which is governed by a jump process with a power-law hazard function have been determined, one can investigate how the maximum log-likelihood of these parameters behaves when applied to simulated data.

6.2.2 Detecting whether a market is in a bubble regime

In the following sections, a strategy to give a measure of the likelihood that a particular market is in a bubble regime is developed, and in doing so areas are identified in observed data where the market is moving in and out of periods more likely to be exhibiting superexponential growth. The strategy is developed over

the next few subsections and the implementation details are given in §6.2.3.

The first part of the strategy is to simulate normal and superexponential markets N times by building models based on the theory in §6.2.1. Let $m = 1, \ldots, N$ enumerate the models. Then each model is determined by a choice of parameters μ_m, σ_m, ν_m, β_m and $t_{c,m}$. The model is then simulated by random variation of parameters, including the initial time t_1, keeping the final time fixed. Discrete-time simulation is used with the n equally spaced times (market days), as described below. The aim is to use the simulations to obtain data from which an implied representative value ν' can be obtained by a maximum log-likelihood estimate. It is the distribution of this representative value of ν that is used to distinguish regimes of normal and superexponential growth.

Construction of the simulated data

In order to look at the distributions of ν as calculated in both normal and superexponential regimes, a large number of sets, say N, of simulated data are constructed for both market types, allowing one to determine whether one should expect to be able to detect changes in such distributions. The simulated data for the mth time-series is constructed such that each data point is given by

$$r_i = x_i - x_{i-1} = \left(\mu_m - \nu_m (t_{c,m} - t_i)^{\beta_m}\right) \Delta + \sigma_m \sqrt{\Delta} N_i(0,1) \tag{6.33}$$

where $i = 1, \ldots, n$, and n is the number of trading-days worth of data in the mth time series (noting that although n varies, this notation is kept for simplicity), $r_1 = 0$, Δ is one trading-day measured in years, and the parameters indexed by m are chosen such that each simulated time-series is representative of the market in question. Here $N_i(0,1)$ is a standard normal distribution. In what follows this notation is used, sometimes indexed, to denote independent normal distributions.

Equation (6.33), which is equivalent to the differences r_i being independently distributed with probability density function

$$F(r_i, t_i) = \frac{1}{\sqrt{2\pi\sigma_{t_i}^2 \Delta}} \exp\left[-\frac{(r_i - \mu_{t_i}\Delta)^2}{2\sigma_{t_i}^2 \Delta}\right], \tag{6.34}$$

is derived from (6.18) and (6.24). In what follows σ is restricted to a constant value to reduce the complexity of the maximum log-likelihood calculations in §6.2.2.

It should be noted that $-1 < \beta_m < 0$ so that one may take $\beta_m \sim U(-1, 0)$, a uniform distribution on $(-1, 0)$. As for t_c, it is known that it should occur after the time t_n. Therefore, $t_c \sim U(t_n, t_n + \gamma)$ where γ is a short time, as described below in §6.2.2.

The distributions for μ_m and σ_m maybe be taken from the observed data of the target market as described in the following paragraph. This leaves one to consider how to sample ν_m. It is not immediately apparent how these values should be distributed, but a closer look at (6.32) can inform the right direction to follow. The distribution of ν_m is discussed at the end of this subsection.

Estimating distributions for μ_m and σ_m Since the aim is to create realistic simulated data for the particular market being the focus of the thesis' investigations, realistic samples of both the drift parameter, μ_m, and its standard deviation, σ_m must be derived such that they may to be used in building the model.

This section takes, as an example, the S&P500 index, considering market close data from January 1950 to June 2018, and calculates the daily mean of the log-price return, $\mu_m\Delta$, and its standard deviation, $\sigma_m\sqrt{\Delta}$ for a large sample $(10,000)$ of randomly chosen time periods (in years) of lengths drawn from $U(0, 10)$.

As can be seen from Fig. 6.5, approximate normal distributions have been fitted to these sample data for $\mu_m\Delta$ and $\log(\sigma_m\sqrt{\Delta})$ and in Fig. 6.6 a simulated joint distribution of the daily mean of the log-price returns has been derived. This joint distribution is used in the construction of the simulated data on which is based the subsequent analysis of this particular time series, as described in §6.2.3.

It is important to note that in this analysis no attempt has been made to accurately model the distribution of the daily means and standard deviations; rather the aim has been to capture the approximate range and frequency of

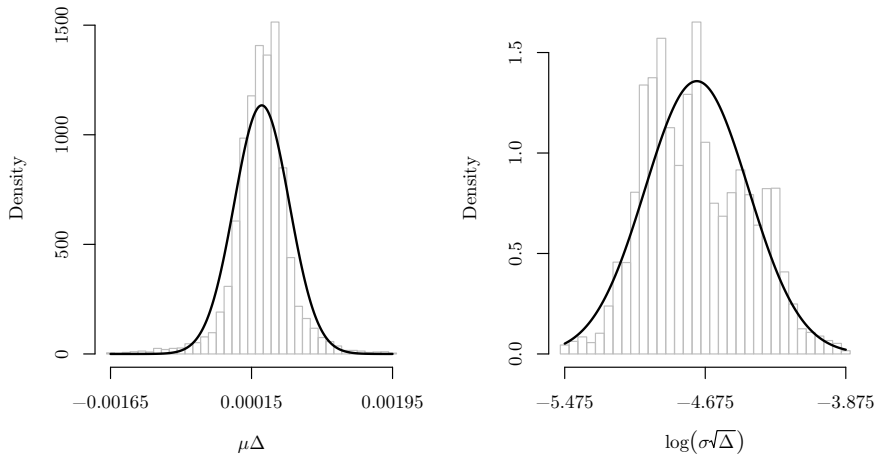


Figure 6.5: Distribution of daily mean (not annualised) and the logarithm of their standard deviations derived from the S&P500 index log-price returns from January 1950 to June 2018. 10,000 sample time-windows were take with spans up to 10 years drawn randomly from $U(0, 10)$. The returns and logarithm of their standard deviations have been fitted with normal distributions. As can be seen, the fits are fair and sufficient for the purposes of building market simulations which are at least representative of historical reality.

values observed in this particular market.

Estimating a realistic distribution of ν_m In order to decide upon a realistic distribution of ν_m to be used in generating our simulated data, one must first be clear on which distributions are reasonable for both B and κ. First, looking at (6.32), it is known that at time $t = t_c$ the probability that a crash has not occurred up until this point is given by

$$P(Y_{t_c}) = \exp\left[-\frac{B}{1+\beta}(t_c - t_1)^{1+\beta}\right] . \tag{6.35}$$

Since β is chosen on $\beta \sim U(-1, 0)$, and the random variable $\tilde{p}$ is defined as $\tilde{p} = P(Y_{t_c})$ (where one must take $\tilde{p} \sim U(0, 1)$ since $P(Y_{t_c})$ is a measure of probability), the probability density function of B can be found by considering the relationship between β and B given by

$$B = -(\beta + 1)\theta^{-(\beta+1)} \log \tilde{p} \, , \tag{6.36}$$

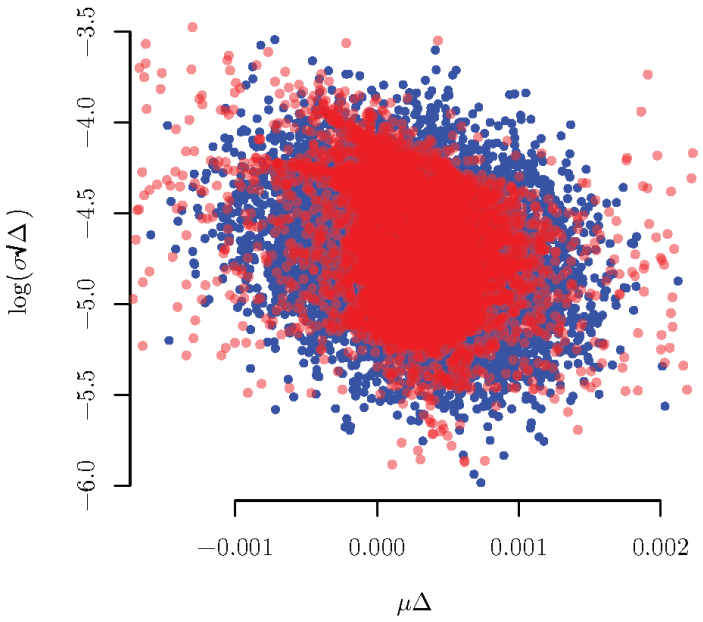


Figure 6.6: Joint distributions of daily $\mu\Delta$ and $\log(\sigma\sqrt{\Delta})$ for the S&P500 index. In red are $10,000$ points of observed data taken from the log-price returns of the index between January 1950 and June 2018, and in blue is the fitted bi-variate normal distribution used for simulating data series using the model given in (6.33). There is an amount of negative correlation between the two parameters, given by a correlation coefficient of -0.25.

where $\theta = t_c - t_1$.

First, consider the function $g(\theta, \beta) = (\beta + 1)\theta^{-(\beta+1)}$ and its partial derivative with respect to β,

$$\frac{\partial g}{\partial \beta} = \theta^{-(\beta+1)}(1 - (\beta + 1)\log\theta), \quad \theta > 0 \tag{6.37}$$

It can be seen that the function $g(\theta, \beta)$ has a unique turning point where $\theta = \exp[1/(\beta + 1)]$. However, since $-1 \leq \beta \leq 0$, a zero of $\partial g/\partial \beta$ only exists when $\theta \geq e$, and for $0 < \theta \leq e$ the function $g(\theta, \beta)$ is monotonically increasing as a function of β.

Now, consider that the random variable B takes a particular value $B = b$ where $b \geq 0$. By rearranging (6.36) and writing $\tilde{p} = p_b(\beta)$ when $B = b$, it can be seen that

$$p_b(\beta) = \exp\left(-\frac{b}{g(\theta, \beta)}\right). \tag{6.38}$$

Since $g(\theta, \beta)$ increases monotonically as a function of β for $0 < \theta \leq e$, this is also true for $p_b(\beta)$. Furthermore, since there is a unique turning point of $g(\theta, \beta)$ while $\theta > e$ this is similarly true for $p_b(\beta)$. These two cases are shown in Fig. 6.7(a) and Fig. 6.7(b) respectively.

Now, since $B = -g(\theta, \beta)\log\tilde{p}$, for a fixed value of β, in the two cases $0 < \theta \leq e$ and $\theta > e$, B is a monotonically decreasing function of $\tilde{p}$. Therefore, it can be seen that the cumulative distribution function of B, $F_B(b) = P(B \leq b)$, can be expressed as

$$P(B \leq b) = 1 - P(B > b) = 1 - \int_{-1}^{0} p_b(\beta)d\beta\,. \tag{6.39}$$

Therefore the probability density function of B can then be found by

$$\begin{aligned} f_B(b) &= -\frac{d}{db}\int_{-1}^{0} p_b(\beta)d\beta \\ &= \int_0^1 \frac{1}{(1-\rho)\theta^{\rho-1}} \exp\left(-\frac{b}{(1-\rho)\theta^{\rho-1}}\right)d\rho \end{aligned} \tag{6.40}$$

where $\rho = -\beta$. To simplify matters a little further for the purposes of constructing the simulated data on which to test the bubble detection method, it is assumed

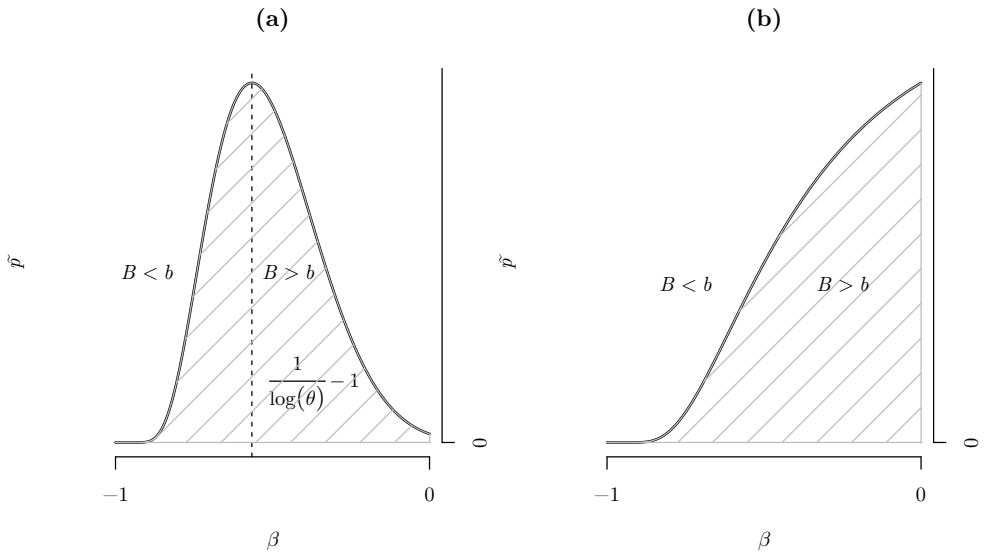


Figure 6.7: For a given value of b, the relationship between $\tilde{p}$ and β is shown where $B = b$. There are two cases, (a) $\theta > e$, and (b) $0 < \theta \leq e$, where θ is the time between the start of the observed data and the critical time t_c. Since $0 \leq \tilde{p} \leq 1$, in both cases the probability that $B < b$ is the area above the curve bounded by $\tilde{p} \leq 1$. Deriving an expression for this area gives the probability density function for B for a given value of θ.

that the time between the critical time, t_c, and the last date in the observed data is small compared to the time span of the data set so that the probability density function becomes

$$f_B(b) = \int_0^1 \frac{1}{(1-\rho)\hat{\theta}^{\rho-1}} \exp\left(-\frac{b}{(1-\rho)\hat{\theta}^{\rho-1}}\right) d\rho \tag{6.41}$$

where $\hat{\theta}$ is the time span of the observed data.

The distribution of the "expected" percentage fall in the market upon the occurrence of a crash, κ, is difficult to determine since market crashes are, on the one hand, rare and, on the other, not subject to strict definition. For the purposes of this thesis, κ is modelled by a uniformly distributed random variable $\kappa \sim U(0, 0.75)$. Other choices of upper limit are certainly possible, and 0.75 may appear somewhat high since a price fall of 75% is unlikely to occur in practice. However since the value of κ is an expectation of draw-down if indeed a crash does happen, one should not necessarily take it as being representative of draw-downs

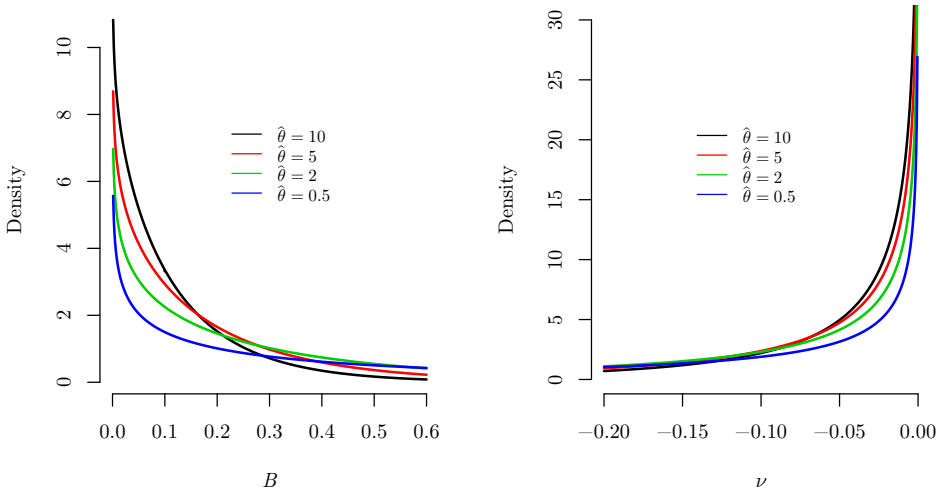


Figure 6.8: Distribution of B on the left and, on the right, the distribution of ν given from (6.41) and (6.42) respectively, for varying values of the observed data time span, $\hat{\theta}$, assuming that the difference between the time t_c and the last date of the observed data is small compared to the total time span. The distribution of ν further assumes that the fraction of the market fall in the event of a crash is a uniform random variable $\kappa \sim U(0, 0.75)$.

which have happened in the real world.

The distribution of the random variable $\nu = B\log(1-\kappa)$ can be written as

$$f_\nu(\nu) = -\int_0^{\kappa_+} \frac{1}{\kappa_+ \log(1-\kappa)} f_B\left(\frac{\nu}{\log(1-\kappa)}\right) d\kappa \tag{6.42}$$

where $\kappa_+ = 0.75$. The probability density functions f_B and f_ν are shown in Fig. 6.8. One can obtain the the parameter ν_m by sampling from the distribution f_ν.

Maximum log-likelihood analysis of the power-law hazard function

Now that a model has been built which can generate simulated data $\mathbf{x} = (x_1, \ldots, x_n)$ via (6.33) it is now described how these data, and in particular, the log returns $\mathbf{r} = (r_1, \ldots, r_n)$, given by $r_1 = 0$, and $r_i = x_i - x_{i-1}$, for $i = 1, \ldots, n$, may be used to obtain maximum log-likelihood estimates for ν.

Equation (6.34) describes the probability density function of the log-price returns as normally distributed parametrised by time-dependent drift and variance terms. Now, given a set of observed asset log-price returns, model parameters

may be estimated using maximum log-likelihood methods; however, to simplify the analysis and allow the parameter ν to be found explicitly, the variance, σ_t^2, is held as a constant such that $\sigma_t^2 = \sigma^2$ giving the simplified probability density function, $\tilde{F}(r,t)$, as

$$\tilde{F}(r_i,t_i) = \frac{1}{\sqrt{2\pi\sigma^2\Delta}} \exp\left[-\frac{(r_i - \Delta(\mu - \nu(t_c - t_i)^\beta))^2}{2\sigma^2\Delta}\right] \tag{6.43}$$

so that

$$\log\left(\tilde{F}(r_i,t_i)\right) = -\frac{1}{2}\log(2\pi\Delta) - \log\sigma - \frac{(r_i - \Delta(\mu - \nu(t_c - t_i)^\beta))^2}{2\sigma^2\Delta} \tag{6.44}$$

Therefore, given a series of log-price returns, $\mathbf{r} = (r_1, r_2, \ldots, r_n)$ and times $\mathbf{t} = (t_1, t_2, \ldots, t_n)$, the log-likelihood function of the parameters is given by

$$\log\mathcal{L}(\mu,\sigma,\nu,t_c,\beta;\mathbf{r}) = -\frac{n}{2}\log(2\pi\Delta) - n\log\sigma - \sum_{i=1}^{n}\frac{(r_i - \Delta(\mu - \nu(t_c - t_i)^\beta))^2}{2\sigma^2\Delta} \tag{6.45}$$

Given this simplified equation, it is possible to find the maximum log-likelihood explicitly for the parameters, μ, σ, and ν. These have been labelled as the linear parameters, with t_c and β labelled as the nonlinear parameters, and expressions are derived to find the value of each parameter that maximises the log-likelihood function, and which are, therefore, the most likely true values which generate the observed data.

Linear parameters. Starting with the parameters μ and ν, the values, μ' and ν', which maximise the log-likelihood function are found by forming the partial derivatives

$$\begin{aligned}
\frac{\partial}{\partial\nu}\log\mathcal{L} &= -\frac{1}{\sigma^2}\sum_{i=1}^{n}(t_c - t_i)^\beta\left(r_i - \Delta(\mu - \nu(t_c - t_i)^\beta)\right) = 0 \\
\frac{\partial}{\partial\mu}\log\mathcal{L} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(r_i - \Delta(\mu - \nu(t_c - t_i)^\beta)\right) = 0
\end{aligned} \tag{6.46}$$

and, as such, are given uniquely by

$$\mu' = \frac{c_1 c_2 - n c_3 \bar{\mathbf{r}}}{\Delta c_4}, \quad \nu' = \frac{n(c_2 - c_1 \bar{\mathbf{r}})}{\Delta c_4} \tag{6.47}$$

where

$$c_1 = \sum_{i=1}^{n} (t_c - t_i)^{\beta}, \quad c_2 = \sum_{i=1}^{n} (t_c - t_i)^{\beta} r_i, \quad c_3 = \sum_{i=1}^{n} (t_c - t_i)^{2\beta} \tag{6.48}$$

and

$$\bar{\mathbf{r}} = \frac{1}{n} \sum_{i=1}^{n} r_i, \quad c_4 = c_1^2 - n c_3 \,. \tag{6.49}$$

Finally, the partial derivative of the log-likelihood function with respect to σ is

$$\frac{\partial}{\partial \sigma} \log \mathcal{L} = -\frac{n}{\sigma} + \frac{1}{\sigma^3 \Delta} \sum_{i=1}^{n} \left(r_i - \Delta(\mu - \nu (t_c - t_i)^{\beta}) \right)^2 = 0 \,, \tag{6.50}$$

so that σ'^2 can be explicitly expressed as

$$\sigma'^2 = \frac{1}{n\Delta} \sum_{i=1}^{n} \left(r_i - \Delta(\mu' - \nu' (t_c - t_i)^{\beta}) \right)^2 \,. \tag{6.51}$$

Since the value which maximises the log-likelihood function for each of the linear parameters may be determined uniquely for given values of t_c and β, it is necessary to first determine the maximising values t_c' and β'.

Non-linear parameters. By differentiating with respect to t_c and β, the non-linear parameters must satisfy

$$\frac{\nu' \beta}{\sigma'^2} \sum_{i=1}^{n} \frac{1}{t_c - t_i} \left(\Delta \left(\mu' (t_c - t_i)^{\beta} - \nu' (t_c - t_i)^{2\beta} \right) - r_i (t_c - t_i)^{\beta} \right) = 0 \,, \tag{6.52}$$

and

$$\frac{\nu'}{\sigma'^2} \sum_{i=1}^{n} \log(t_c - t_i) \left(\Delta \left(\mu' (t_c - t_i)^{\beta} - \nu' (t_c - t_i)^{2\beta} \right) - r_i (t_c - t_i)^{\beta} \right) = 0 \,. \tag{6.53}$$

It can be seen that it is not possible to determine these nonlinear parameters explicitly. However, something is known about their constraints. First, it is necessary that $-1 < \beta < 0$ such that the probability of a crash occurring, on the condition that it has not already happened, remains finite. As for t_c, we know that it should occur after the time t_n, for the same reason. Furthermore, if t_c is very far away from t, it is unlikely that the effect of the power-law would be detectable. Therefore, one can sample values of $\beta \sim U(-1, 0)$ and $t_c \sim U(t, t+\gamma)$ where γ is a short time interval, and derive distributions for μ' and σ', but more importantly for ν'. In what follows, it is assumed that $\gamma = 0.5$ year.

Maximum log-likelihood analysis of ν in simulated data

ν' distribution Equations (6.47) give μ' and ν' from the maximum log-likelihood estimates, given a simulated set of log returns $\mathbf{r} = (r_1, r_2, \ldots, r_n)$ observed at times $\mathbf{t} = (t_1, t_2, \ldots, t_n)$. Recall from the beginning of §6.2.2 that N models are built with the mth model given by (6.33). In what follows, the returns data $r_{i,m}$, $i = 1, \ldots, n$ are taken for the mth model to be input into a simulation where n, t_c and β are random variables which are independent of the model chosen. Therefore the maximum log-likelihood estimates must be modified slightly to take into account the difference between the model generating the log returns data and that being used in the simulation. Specifically, (6.45) is used to obtain the maximum log-likelihood estimates for ν (and μ), with $r_i = r_{i,m}$, $i = 1, \ldots, n$, where n is the (randomly varying) length of the time series.

Recall from (6.46) that a time series of log-price returns $\mathbf{r}$ observed at times $\mathbf{t}$, the values of ν' and μ' that maximise their respective log-likelihoods are given by

$$\begin{aligned} \nu' &= \frac{\mu' \sum_{i=1}^{n} (t_c - t_i)^{\beta} - \Delta^{-1} \sum_{i=1}^{n} (t_c - t_i)^{\beta} r_i}{\sum_{i=1}^{n} (t_c - t_i)^{2\beta}} \\ \mu' &= \frac{\nu'}{n} \sum_{i=1}^{n} (t_c - t_i)^{\beta} + \Delta^{-1} \frac{1}{n} \sum_{i=1}^{n} r_i \end{aligned} \tag{6.54}$$

Setting $r_i = r_{i,m}$, the sum of the simulated log-price returns is given by

$$\begin{aligned}\sum_{i=1}^{n} r_{i,m} &= \sum_{i=1}^{n}\Big((\mu_m - \nu_m(t_{c,m}-t_i)^{\beta_m})\Delta + \sigma_m\sqrt{\Delta}N_i(0,1)\Big) \\ &= n\mu_m\Delta - \nu_m\Delta k_1 + \sum_{i=1}^{n}\sigma_m\sqrt{\Delta}N_i(0,1)\end{aligned} \tag{6.55}$$

where $k_1 = \sum_{i=1}^{n}(t_{c,m}-t_i)^{\beta_m}$. Recalling the notation in (6.48), and substituting into equation (6.54), one has

$$\begin{aligned}\mu' &= \frac{\nu' c_1}{n} + \frac{1}{n}\left(n\mu_m - \nu_m k_1 + \sum_{i=1}^{n}\frac{\sigma_m}{\sqrt{\Delta}}N_i(0,1)\right) \\ &= \mu_m + \frac{1}{n}(c_1\nu' - k_1\nu_m) + S_1\end{aligned} \tag{6.56}$$

where $S_1 = \frac{1}{n}\sum_{i=1}^{n}\frac{\sigma_m}{\sqrt{\Delta}}N_i(0,1)$. Furthermore,

$$\begin{aligned}\nu' &= \frac{1}{c_3}\left(\mu' c_1 - \mu_m c_1 + \nu_m k_3 - \sum_{i=1}^{n}\frac{\sigma_m}{\sqrt{\Delta}}(t_c - t_i)^{\beta}N_i(0,1)\right) \\ &= \frac{k_3}{c_3}\nu_m + \frac{c_1}{c_3}(\mu' - \mu_m) - S_2\end{aligned} \tag{6.57}$$

where $S_2 = \dfrac{1}{c_3}\displaystyle\sum_{i=1}^{n}\frac{\sigma_m}{\sqrt{\Delta}}(t_c - t_i)^{\beta}N_i(0,1)$ and $k_3 = \sum_{i=1}^{n}(t_c - t_i)^{\beta}(t_{c,m} - t_i)^{\beta_m}$. Now, solving for ν',

$$\nu' = \frac{1}{nc_3 - c_1^2}\left((k_3 n - c_1 k_1)\nu_m - n(c_3 S_2 - c_1 S_1)\right). \tag{6.58}$$

However

$$\begin{aligned}c_3 S_2 - c_1 S_1 &= \sum_{i=1}^{n}\left(\frac{\sigma_m}{\sqrt{\Delta}}(t_c - t_i)^{\beta} - \frac{c_1\sigma_m}{\sqrt{\Delta}n}\right)N_i(0,1) \\ &= N_m(0,1)\,\frac{\sigma_m}{\sqrt{\Delta}}\sqrt{\left(c_3 - \frac{c_1^2}{n}\right)}\end{aligned} \tag{6.59}$$

where $N_m(0,1)$ is a single normally distributed random variable. Therefore

$$\nu' = \frac{1}{nc_3 - c_1^2}\left((k_3 n - c_1 k_1)\nu_m - nN_m\left(0,1\right)\sigma_m\sqrt{\frac{1}{\Delta}\left(c_3 - \frac{c_1^2}{n}\right)}\right). \tag{6.60}$$

Furthermore, for analysis of simulated data that follow a Brownian motion stochastic differential equation, ν_m is simply set as $\nu_m = 0$ and (6.60) becomes

$$\nu' = N_m\left(0,1\right)\frac{\sigma_m}{\sqrt{\Delta\left(c_3 - \frac{c_1^2}{n}\right)}}. \tag{6.61}$$

Note that, although $\nu_m \leq 0$, there is no such restriction in the distribution of ν', because random variation may cause the maximum log-likelihood estimate to correspond to negative growth, even when $\nu_m < 0$. In fact, as is to be expected, the distribution in the case of superexponential growth falls principally on negative values of ν', but for $\nu_m = 0$, the distribution is symmetric about $\nu' = 0$.

Approximation for large N

Given the large number, N, of models needed to derive a properly representative picture of the distributions of ν', it is important to approximate the sums in (6.60) and (6.61) to reduce the computing time required.

Replacing the sums with integral approximations, it is straightforward to obtain

$$c_1 = \sum_{i=1}^{n}(t_c - t_i)^{\beta} \approx \hat{c}_1 = \frac{1}{\Delta}\frac{(t_c - t_1)^{\beta+1} - (t_c - t_n)^{\beta+1}}{\beta + 1} \tag{6.62}$$

$$c_3 = \sum_{i=1}^{n}(t_c - t_i)^{2\beta} \approx \hat{c}_3 = \frac{1}{\Delta}\frac{(t_c - t_1)^{2\beta+1} - (t_c - t_n)^{2\beta+1}}{2\beta + 1} \tag{6.63}$$

$$k_1 = \sum_{i=1}^{n}(t_{c,m} - t_i)^{\beta_m} \approx \hat{k}_1 = \frac{1}{\Delta}\frac{(t_{c,m} - t_1)^{\beta_m+1} - (t_{c,m} - t_n)^{\beta_m+1}}{\beta_m + 1}. \tag{6.64}$$

An approximation to k_3, of order $a \geq 0$, is

$$\hat{k}_3 = \frac{1}{\Delta} \begin{cases} \displaystyle\sum_{j=0}^{a} \binom{\beta_m}{j} (t_{c,m} - t_c)^j \frac{(t_c - t_1)^{\alpha_m - j + 1} - (t_c - t_n)^{\alpha_m - j + 1}}{\alpha_m - j + 1} & t_{c,m} < t_c \\ \dfrac{(t_c - t_1)^{\alpha_m + 1} - (t_c - t_n)^{\alpha_m + 1}}{\alpha_m + 1} & t_{c,m} = t_c \\ \displaystyle\sum_{j=0}^{a} \binom{\beta}{j} (t_c - t_{c,m})^j \frac{(t_{c,m} - t_1)^{\alpha_m - j + 1} - (t_{c,m} - t_n)^{\alpha_m - j + 1}}{\alpha_m - j + 1} & t_{c,m} > t_c \end{cases}$$

where $\alpha_m = \beta + \beta_m$.

Using these approximations, which work remarkably well in practice, the expressions for ν' in (6.60) and (6.61) become

$$\nu' = \frac{1}{n\hat{c}_3 - \hat{c}_1^2} \left((\hat{k}_3 n - \hat{c}_1 \hat{k}_1)\nu_m - nN_m(0,1)\,\sigma_m \sqrt{\frac{1}{\Delta}\left(\hat{c}_3 - \frac{\hat{c}_1^2}{n}\right)} \right). \tag{6.65}$$

and

$$\nu' = N_m(0,1) \frac{\sigma_m}{\sqrt{\Delta\left(\hat{c}_3 - \frac{\hat{c}_1^2}{n}\right)}}. \tag{6.66}$$

Example fitted distributions of ν' most likely value Now it is necessary briefly take stock of the development thus far. For each of $m = 1, \ldots, N$ there are randomly selected parameters μ_m, σ_m, ν_m, β_m and $t_{c,m}$. For each of these parameter choices, the model is simulated by taking a large random sample of times $t_1, \ldots, t_n$, with t_n fixed but n, and hence t_1, varying randomly, and parameters $\beta \sim U(-1, 0)$ and $t_c \sim U(t_n, t_n + \gamma)$, so that the returns $r_1, \ldots, r_n$ are obtained from (6.33).

For each of these simulated data series, one can find distributions of ν' given sufficient samples drawn for the controlling random variables, β and t_c. However, since the MLEs in (6.47) are highly biased, in that the expected value of ν' may diverge from the simulated parameter value, the most likely value of these distributions, rather than the expected value, is taken as an estimate of the parameter value. Now an investigation is made regarding how the most likely values of these distributions are themselves distributed over many thousands of

simulated data series. The most likely value of the distribution of ν' is defined as $\bar{\nu}$.

As an example, Fig. 6.9 shows the resulting distribution for $\bar{\nu}$ for the normal and superexponential growth cases taken from sample data drawn from the S&P500 index over the past 60 years, as used in §6.2.3. In this figure, the distributions are overlaid with, in both cases, the generalised/skew forms of the Cauchy, normal, Laplace and logistic distributions. These fits were obtained by the `fitdistrplus` package as used in the statistical computing platform, R. In Table 6.1 summary statistics are given for all fitted distributions.

	Normal growth			
Measure	Logistic	Laplace	Cauchy	Normal
AIC	9954	10151	11324	10142
BIC	9974	10171	11344	10162
KS p-value	0.0059	0.0366	0.0652	0.0275
KS test	not rejected	rejected	rejected	rejected
	Superexponential growth			
Measure	Logistic	Laplace	Cauchy	Normal
AIC	11448	11504	12473	12204
BIC	11467	11524	12492	12224
KS p-value	0.0188	0.0321	0.0585	0.0695
KS test	not rejected	rejected	rejected	rejected

Table 6.1: Simulated data $\bar{\nu}$ distribution fitting statistics, showing the generalised/skew logistic distributions to be the best fits for both normal and superexponential growth.

By all standard measures, the generalised logistic distributions are the best fits, with none of the other distributions not being rejected at the 5% level for the Kolmogorov-Smirov test. Although these fits are clearly not exact around the modal value, there is remarkably good agreement with 5000 data points. In the normal and superexponential cases the fits cannot be rejected in a Kolmogorov-Smirnov (KS) test at the 5% level.

A better fit may be possible using a distribution that truncates the (generalised) logistic distribution, but a more refined fit is unlikely to improve the application to financial indices, as the distribution of $\mu_m\Delta$ and $\log(\sigma_m\sqrt{\Delta})$ is

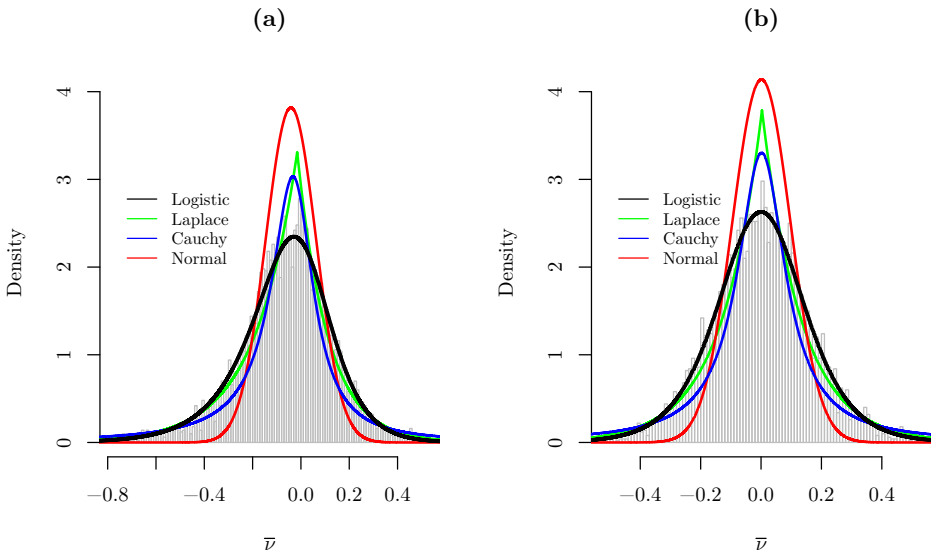


Figure 6.9: Distributions of $\bar{\nu}$, given the experimental simulations of the (a) superexponential and (b) normal regimes. Having trialled a number of candidate distributions against the experimental data, a generalised logistic distribution appears to fit the data well for both the normal and superexponential governing regimes.

only approximated by the bivariate normal distribution shown Fig. 6.6

Threshold value of $\bar{\nu}$ In order to determine whether a particular time series is in a normal or superexponential growth a Bayesian approach to hypothesis testing is taken such that there are two potential hypotheses: H_1 being the case in which the most likely value, $\bar{\nu}$, indicates a superexponential governing regime, and H_0, the case in which $\bar{\nu}$ indicates the market is following a normal growth period. H_1 is accepted and H_0 is rejected when $P(H_1|\bar{\nu}) > P(H_0|\bar{\nu})$. Since $P(H_i|\bar{\nu}) = f_{H_i}(\bar{\nu})P(H_i)/P(\bar{\nu})$, where $f_{H_i}(\bar{\nu})$ is the probability density function of $\bar{\nu}$ given the underlying hypothesis, H_1 is accepted when

$$f_{H_1}(\bar{\nu}) > f_{H_0}(\bar{\nu})\frac{P(H_0)}{P(H_1)}\,. \tag{6.67}$$

Assuming that the probability of the governing regime being superexponential $P(H_1) = p$, the threshold is the value of $\bar{\nu}$ for which

$$f_{H_1}(\bar{\nu}) = f_{H_0}(\bar{\nu})\frac{1-p}{p}\,. \tag{6.68}$$

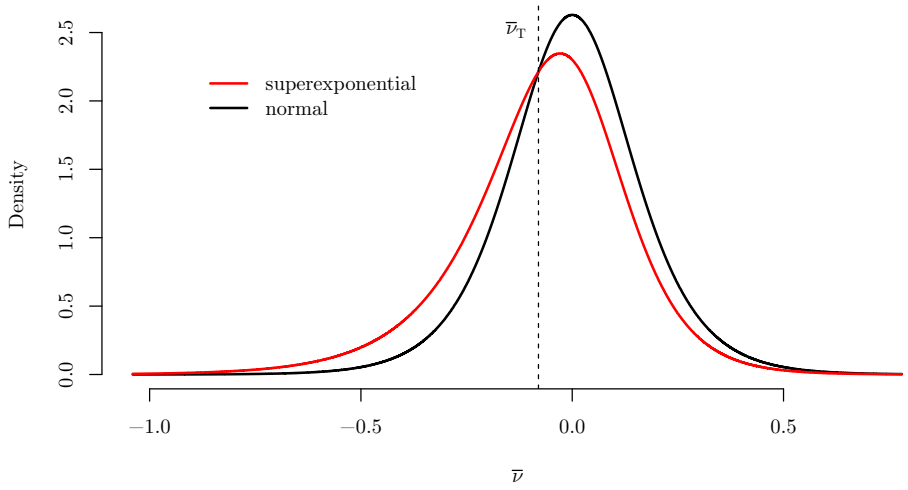


Figure 6.10: A comparison of the logistic-fitted distributions of the superexponential and normal regimes. The threshold value $\bar{\nu}_T$ is interpreted as the lower bound on the value of $\bar{\nu}$ at which one should expect the governing regime to be normal. This makes the assumption that the probability of the regime being in one of the two possible regimes is 0.5.

There is no prior knowledge of the value of p, and a value of $p = 0.5$ is assumed such that the threshold value, $\bar{\nu}_T$ is given by the solution of $f_{H_1}(\bar{\nu}) = f_{H_0}(\bar{\nu})$. Therefore, when applying this threshold to market data, by calculating the most likely value of the log-likelihood maximising distribution of ν' for a particular time period, one can simply say that this regime is more probably in a normal growth phase if $\bar{\nu} > \bar{\nu}_T$ where $\bar{\nu}_T$ is the threshold value calculated from the simulations of H_1 and H_0 derived from the statistical parameters of the market in question.

6.2.3 Applying $\bar{\nu}_T$ log-likelihood threshold to the S&P500 index

The methodology developed in the preceding sections has been used to examine the S&P500 index from January 1960 up to June 2018, to determine whether the market is exhibiting superexponential growth at any particular point in time. The precise algorithm for the calculation of the distribution of $\bar{\nu}$ and of the threshold value $\bar{\nu}_T$ is somewhat involved and is described below as a sequence of steps.

Some of these steps have been described earlier by way of example of the theory, but are included here for completeness.

Gather market data for simulations. Taking the S&P500 index from 1950, the (daily) log-returns were calculated. Random time spans of up to 10 years with random starting points (consistent with the period January 1950 – June 2018) were sampled and on each of these random intervals the mean daily log return $\mu\Delta$ and its standard deviation $\sigma\sqrt{\Delta}$ were calculated and fitted to a bi-variate normal distribution. In this study, 10,000 random intervals were taken. In Fig. 6.6, a comparison of the data and fitted distribution is illustrated, showing fair agreement between the model and the data.

Simulate market regimes to find $\bar{\nu}$ distribution. The second stage is to simulate log-returns using the stochastic models for normal and superexponential growth. Taking the arbitrary, but convenient 10-year period 1980 – 1990, a stochastic model was built, and simulated data for each of the two cases were generated, as described below.

Normal growth. 5000 simulations were built for normal growth by randomly choosing μ_m, σ_m via the bi-variate normal distribution given above, $t_{c,m}$ from the uniform distribution $U(1990, 1990.5)$ and β_m from the uniform distribution $U(-1, 0)$ and choosing $\nu_m = 0$.

Then, for each model, 5000 values of ν' were obtained from the maximum likelihood estimates given in (6.61) with randomly sampling values of $t_c \sim U(1990, 1990.5)$ and $\beta \sim U(-1, 0)$ taking a time interval up to 1990 from a starting point drawn randomly from the uniform distribution $U(1980, 1990)$. From 5000 values of ν' a most likely value $\bar{\nu}_m$ was determined by using the `modeest` package in R.

Superexponential growth. For superexponential growth, the procedure was similar to that of normal growth with the crucial difference that ν_m were chosen randomly from the distribution (6.42). As before, the model was then simulated 5000 times and a most likely value $\bar{\nu}_m$ was obtained from the 5000 values of ν' calculated from the maximum likelihood estimates (6.60).

Fit generalised logistic distributions and find $\bar{\nu}_T$. In both cases, the distributions of $\{\bar{\nu}_m : m = 1, \ldots, 5000\}$ were well described by generalised logistic distributions as shown in Fig. 6.9. By examining these distributions, $\bar{\nu}_T$ was calculated. This is a threshold value which is specifically relevant for the S&P500 index.

Find most likely value of $\bar{\nu}$ for each trading day of the observed data. Now, for each trading day of the S&P500 index from 1960 up until the present day the distribution of ν' was calculated by taking 5000 random samples of the parameters t_c and β from the same distributions as for the simulated data and applying (6.47) for randomly chosen time periods of up to ten years.

Identify regime change-points. Finally, for each trading day, the calculated value of $\bar{\nu}$ was compared to the value of $\bar{\nu}_T$. Where $\bar{\nu} \leq \bar{\nu}_T$ it was determined that the market at that point was governed by a superexponential regime.

The results of this process are shown in Fig. 6.11. From a visual inspection, the areas where it is determined that superexpontial growth as the more likely governing regime (as shown in the blue shading) seem to correspond almost exclusively to periods of growth. Encouragingly for the algorithm, superexponential growth is not generally detected in periods of observed negative growth, and there are also long periods of actual growth in the market which is not detected as superexponential. This is what one would expect to find based on the initial assumptions.

On closer examination, it appears that, at least when viewed from an *ex post* perspective, there may be a predictive quality to the algorithm, in that just prior to crashes, steep sell-offs, and longer declines, there are shifts from one regime to the other demonstrated by a well-defined boundary between shaded and subsequently non-shaded areas. These are the change-points that have been sought in the data.

However, as one would expect with observed data, the results are subject to a fair degree of "noise", in that there are periods of superexponential growth inter-

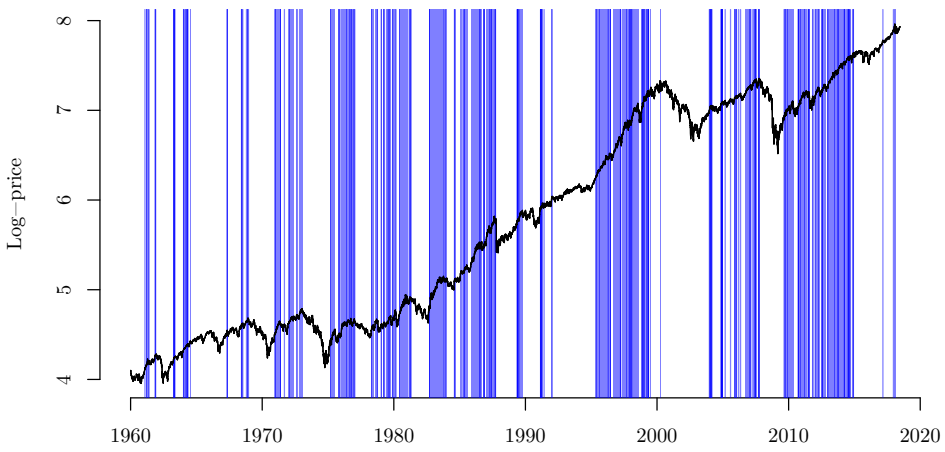


Figure 6.11: Results from applying the threshold value, $\bar{\nu}_T$, found from the fitted distributions of $\bar{\nu}$ calculated from both superexponential and normal market simulations, to the historical daily log-returns of the S&P500 index from January 1960 to June 2018. Areas where the calculated value of $\bar{\nu}$ for the observed data is less than $\bar{\nu}_T$ are shaded in blue. One can see a striking visual correlation between periods of high growth and the blue shaded area, and also a general agreement between the right boundaries of visually significant areas of blue shading and subsequent market downturns.

spersed with brief periods of normal growth. Since it is unlikely that the market governing regime would switch this rapidly, we conclude the algorithm does not capture the granular-level market features, but appears to identify remarkably well the gross features of the market over longer timescales. Note that it appears that longer periods of normal growth are not generally interspersed by shorter periods of superexponential growth.

Focus now turns to some notable historical financial events, and the results from application of the algorithm over shorter timescales. In particular, the following paragraphs look at the 10-year periods preceding the great crash of October 1987, so called Black Monday, the financial crisis of 2007 – 2009, and the 2015 – 2016 stock market sell-off. These are shown in Fig. 6.12 as plots (a), (b) and (c) respectively.

Black Monday. The most notable, and well studied, financial crash of the second half of the last century is the stock market crash of Monday, 19 October

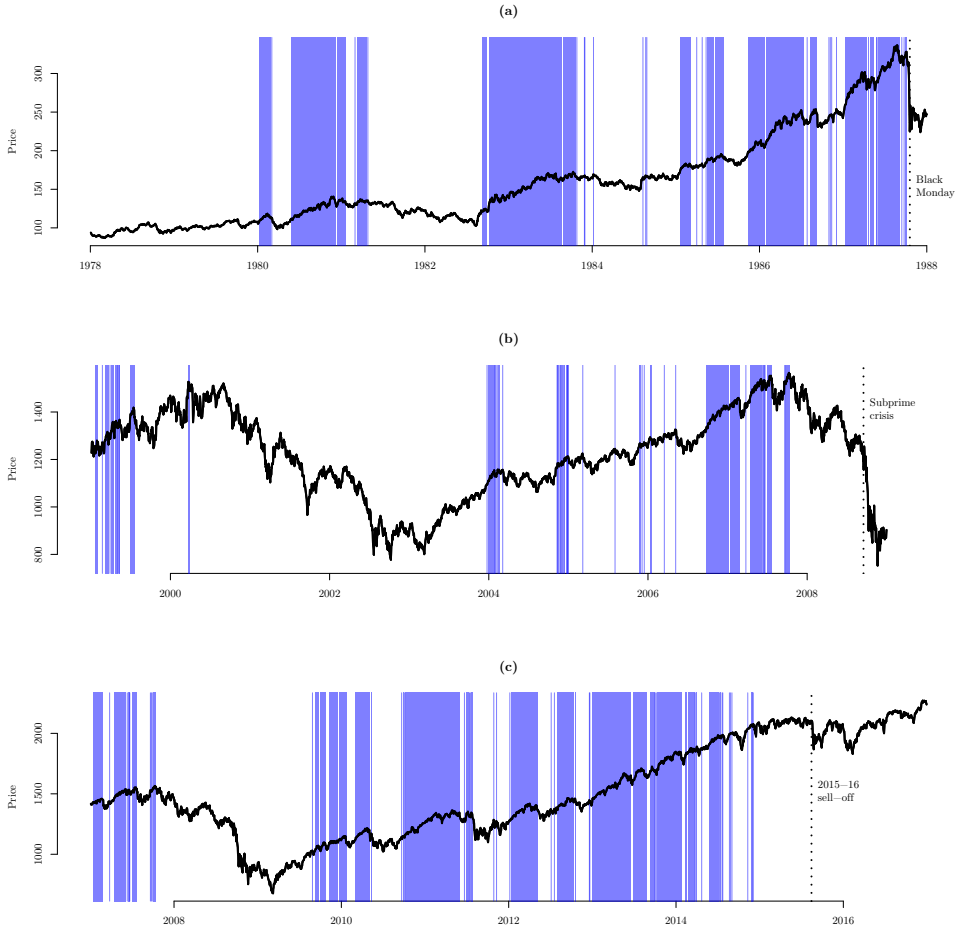


Figure 6.12: Results from applying the threshold value, $\bar{\nu}_T$, found from the fitted distributions of $\bar{\nu}$ for specific periods in the last 60 years of the S&P500 index. (a) Black Monday, October 1987, (b) the financial crisis of 2007 – 2009, and (c) the 2015 – 2016 stock market sell-off. Note that for clarity the price rather than the log price is shown.

1987. There has been much debate over the causes of this crash and the subsequent contagion across the global financial markets (Barro et al. 1989). When viewed from an econophysics perspective, the crash has often been attributed to critical behaviour following a period of superexponential power-law growth decorated with log-periodic oscillations (Sornette et al. 1996 and Feigenbaum & Freund 1996).

By the method presented here, bearing in mind that, for each point ten years' worth of preceding market data is used, superexponential growth modelled by the log-periodic power-law models is detected for the period immediately prior to Black Monday. However, there are also periods in which superexponential growth is not detected. Indeed, each market rally shows superexponential growth, and when the market is in an intermediate downturn, no superexponential growth is found. There is clearly an oscillatory pattern in the market data, and the algorithm picks up this feature. Furthermore, when looked at closely, with the benefit of hindsight, there are fluctuating periods of sub- and super-threshold regions end some days prior to the crash.

Financial crisis of 2007 – 2009. This period was certainly the worst financial crisis of this century and, probably was the most severe since the Great Depression of the 1930s. The US market peaked in October 2007 and there followed a period of steep decline until the global markets suffered a significant draw-down when the fall-out of the subprime mortgage market led to the collapse of large US banking institutions, such as Lehman Brothers, on 15 September 2008. What followed was a global economic downturn, which became known as the Great Recession.

The period 2003 – 2007 was a bull market. However, when one looks at the calculated intervals of superexponential growth in this period, sustained superexponential growth is only seen from late 2006. Contrary to the example of Black Monday, the superexponential regime does not end with a crash following immediately afterwards, but rather a steep downturn continuing for some time prior to a more severe crash.

2015 – 2016 stock market sell-off. Since the end of the Great Recession in 2009, the market followed an almost relentless bull run up to the end of the observed data (June 2018) which has been used in this analysis. This is one of the longest bull markets in history, and, according to the analysis, the majority of it has been governed by a superexponential regime from late 2009 until the beginning of 2015. There does not appear to be the oscillatory patterns of the bull market in the 10 years prior to Black Monday, and, looking at the data from an *ex post* perspective, it is clear that the superexponential period ended with a relatively long period of growth. Here we see an example of a superexponential bubble, which peters out rather than suffering a catastrophic burst. However, the following August the market did suffer large losses which took until the following June to recover.

These examples show that if there is merit in the algorithm's ability to detect periods where the markets are in a bubble regime and experiencing superexponential growth, the transitioning from one regime to the other is not typified by either a catastrophic crash or a market downturn. However, in general there is a good agreement that market downturns are preceded by a regime change, and this appears to be evident on a variety of scales.

6.2.4 Change-point models concluding remarks

In this chapter, a description has been given of two methods that have been investigated to detect discrete scale invariance in observed financial data, using the S&P500 index as a testing ground.

Of these methods, the most promising approach has been found to be derived from the maximum log-likelihood analysis of the log-normal stochastic differential equation with power-law hazard function model of financial markets. This method can be adapted to give on-line indications of whether a market is currently in a bubble regime or not.

It has been seen that the distributions of the parameter $\bar{\nu}$ for the normal and superexponential growth regimes are very well fitted by generalised/skew logis-

tic distributions. Although logistic distributions have been used successfully to model financial assets (Olson & Wu 2013, Tolikas & Gettinby 2009 and Tolikas & Brown 2006), further work is needed to elicit the reason for the logistic distributions in this case, and this research path is not taken in what remains of this thesis. Although not directly relevant to the distribution of $\bar{\nu}$, it is worth noting that the logistic distribution is known to arise in respect of the distribution of the midrange samples taken from a continuous distribution (Gumbel 1944).

In the modelling presented herein a full model of the time series has not been used, but rather the focus has been on the influence of the parameter ν. Indeed, the simulated data with superexponential mean log-returns have been specifically constructed such that the simulations have constant standard deviations in the stochastic terms, and research into these simulated times series has been restricted to the case of constant σ_t. This thesis specifically conjectures that it is the existence, or otherwise, of the hazard function in the drift term alone that can determine whether regime change has occurred. Although this simplification greatly eases the analysis, allowing a closed-form solution to the log-likelihood minimising value of ν, it is to be hoped that future theoretical developments will provide a more complete theory incorporating time-varying σ. Again, this is beyond the scope of research presented here.

In the next chapter, the financial significance of the parameter $\bar{\nu}$ needs further study. Formally, it is the coefficient of the hazard function corresponding to a superexponential financial bubble, but as will be seen later, there is a more intuitive interpretation that can be given in terms of market behaviour under a power law time transform.

The investigations so far have been restricted to the S&P500 index, albeit over 60 years, which constitutes a serious longitudinal study. However, it will be seen how an enhancement of this methodology can lead to computational saving so that one may apply the algorithm to a wide variety of financial time series in a very efficient manner. Importantly the practical utility of the distributions of $\bar{\nu}$ for financial practitioners is shown, including how $\bar{\nu}$ may be used by traders either for short-term trades (both on the index itself and/or derivatives) or for

market timing for longer-term trading strategies. Indeed, it is proposed that in some cases $\bar{\nu}$ may be used as an early-warning for financial crashes and/or the end of a bull market, or indeed the end of periods of anti-bubble. It is hoped that the theory and results presented in the next chapter will be of significant interest to practitioners and theoreticians alike.

Chapter 7

Power-law model for predicting financial markets price momentum

In Chapter 6, the parameter $\bar{\nu}$ was introduced as an indication of whether an observed market was experiencing superexponential growth, and as such exhibiting discrete scale invariance due to the cooperative behaviour of market agents. In order to determine the likelihood of the existence of superexponetial growth, this measure was compared with a threshold value, $\bar{\nu}_T$, which had been determined by comparing simulations of normal growth and superexponetial growth given historical distributions of mean and standard deviation of returns in the chosen market over the entire period of available observations.

The algorithm described in §6.2.3 is a complicated process and, despite the approximations that have been made in §6.2.2, requires a great deal of computations. In the first instance, one must determine the threshold value separating periods of normal growth from superexponential growth, by generating many thousands of sample log-price trajectories given a distribution of means and standard deviations taken over many thousands of observed data windows. Over each of these simulations, one must find the distributions of the calculated values of ν' given sampled values of data window length and the random variables, t_c and β. Once the value of $\bar{\nu}_T$ has been determined, one may then find the calculated

distribution of ν' for any point in the observed data, again for many thousands of sampled values of the critical time, the exponent and the amount of historical data used in such a calculation. One can see that applying this algorithm to a huge cross-sectional study in real-time is quite cumbersome from a computational perspective, and additionally, the reliance on the threshold parameter being the arbiter of the governing regime is unsatisfactory given the assumptions underlying its calculation.

In this chapter, a methodology is investigated that dispenses with the threshold value, $\bar{\nu}_T$, and instead derives a proxy to $\bar{\nu}$ which can be calculated analytically at each point in time, based on a window of observed historical data. This not only removes the computational overhead of the initial simulations required for each market to determine the appropriate threshold, but also removes the need to determine the distribution of ν' by taking numerous samples of t_c and β for each data point. Therefore, an indicative parameter of the governing regime can be calculated for any point in time, for any set of observed data in a computationally inexpensive manner.

Once such a time series has been derived, the quantile of the distribution of the parameter over a set historical time window given by the parameter value on a specific date can be easily determined. This probabilistic measure of the prevailing governing regime is then used to predict market movements on an *ex ante* basis. Results are presented which indicate an interesting predictive quality of the derived parameter when applied to the S&P500 index.

7.1 Interpretation of ν' and $\bar{\nu}$

The parameter ν' described in Chapter 6 is formally the log-likelihood maximising growth of the difference in an asset's log-prices over a fixed period, with the time dimension transformed by a power-law conditional on the critical time, t_c, of an assumed bubble or anti-bubble regime, lying in the region $(t_n, t_n + \epsilon)$ where ϵ is some chosen period of time. This concept is explored more fully below.

In the next section, the following convention is adopted: sets of observed

values are denoted in lower case and random variables in upper case. There are two exceptions to this, namely the random variable for the critical time remains t_c so that consistency with previous notation is preserved, and similarly for the exponent random variable β .

7.1.1 Ordinary least squares by maximum log-likelihood

Assume one has data $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ observed at times $\mathbf{t} = (t_1, t_2, \ldots, t_n)$ and each y_i is independent and distributed normally about a value μ_i with a standard deviation σ. This being the case, the probability density of observing y_i at time t_i is given by

$$P(y_i) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_i-\mu_i)^2}{2\sigma^2}} \tag{7.1}$$

Therefore, the likelihood of observing the entire set of observed data, $\mathbf{y}$, given the parameters is

$$\mathcal{L}(\mathbf{y}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_i-\mu_i)^2}{2\sigma^2}} \tag{7.2}$$

which, taking logarithms, becomes the log-likelihood

$$\ell(\mathbf{y}) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma - \sum_{i=1}^{n} \frac{(y_i-\mu_i)^2}{2\sigma^2}. \tag{7.3}$$

Assuming that the mean of the observed data at any particular point, μ_i, is a linear function of the observation time t_i, such that $\mu_i = \lambda_0 + \lambda_1 t_i$, then, by simple substitution into the above expression, it can be seen that

$$\ell(\mathbf{y}) = -n\log 2\pi - n\log\sigma - \sum_{i=1}^{n} \frac{(y_i-\lambda_0-\lambda_1 t_i)^2}{2\sigma^2}. \tag{7.4}$$

Differentiating with respect to λ_0 and λ_1 gives

$$\begin{aligned} \frac{\partial \ell(\mathbf{y})}{\partial \lambda_0} &= \sum_{i=1}^{n} \frac{y_i-\lambda_0-\lambda_1 t_i}{\sigma^2} \\ \frac{\partial \ell(\mathbf{y})}{\partial \lambda_1} &= \sum_{i=1}^{n} \frac{t_i(y_i-\lambda_0-\lambda_1 t_i)}{\sigma^2}. \end{aligned} \tag{7.5}$$

Equating these partial derivatives to zero forms the simultaneous equations

$$\begin{aligned}\lambda_0 &= \frac{\sum_{i=1}^{n} y_i - \lambda_1 \sum_{i=1}^{n} t_i}{n} \\ \lambda_1 &= \frac{\sum_{i=1}^{n} y_i t_i - \lambda_0 \sum_{i=1}^{n} \bar{t}_i}{\sum_{i=1}^{n} t_i}\end{aligned} \tag{7.6}$$

giving the unique solution for λ_1 as

$$\lambda_1 = \frac{n\left(\sum_{i=1}^{n} y_i t_i - \bar{\mathbf{y}} \sum_{i=1}^{n} t_i\right)}{n \sum_{i=1}^{n} \bar{t}_i^2 - \left(\sum_{i=1}^{n} t_i\right)^2}\,. \tag{7.7}$$

However, the sample covariance of $\mathbf{y}$ and $\mathbf{t}$ is expressed as $\mathrm{Cov}(\mathbf{y}, \mathbf{t}) = \mathrm{E}[\mathbf{yt}] - \mathrm{E}[\mathbf{y}]\,\mathrm{E}[\mathbf{t}]$, which can clearly be written as

$$\begin{aligned}\mathrm{Cov}(\mathbf{y}, \mathbf{t}) &= \frac{1}{n} \sum_{i=1}^{n} y_i t_i - \frac{1}{n^2} \sum_{i=1}^{n} y_i \sum_{i=1}^{n} t_i \\ &= \frac{1}{n} \left(\sum_{i=1}^{n} y_i t_i - \bar{\mathbf{y}} \sum_{i=1}^{n} t_i \right)\end{aligned} \tag{7.8}$$

and similarly the sample variance, $\mathrm{Var}(\mathbf{t})$ is written as

$$\begin{aligned}\mathrm{Var}(\mathbf{t}) &= \mathrm{E}\left[\mathbf{t}^2\right] - \mathrm{E}[\mathbf{t}]^2 \\ &= \frac{1}{n} \sum_{i=1}^{n} t_i^2 - \frac{1}{n^2} \left(\sum_{i=1}^{n} t_i \right)^2 .\end{aligned} \tag{7.9}$$

Therefore, one can see from (7.7) that λ_1 is given by the very simple relationship

$$\lambda_1 = \frac{\mathrm{Cov}(\mathbf{t}, \mathbf{y})}{\mathrm{Var}(\mathbf{t})} \tag{7.10}$$

Here, the population variance and covariance is used as it is expected that in all cases n is large. Since, $\mu_i = \lambda_0 + \lambda_1 t_i$, the log-likelihood maximised slope of the linear relationship between observed data and the time at which it is observed is given by the ratio of the covariance of the data and times, to the variance of the times. As one would expect, the expression for λ_1 is identical in form to the expression derived for ν' in (6.54), except that in this case the set of observation times, $\mathbf{t}$, are transformed by $(t_c - t_i)^\beta$. This suggests that there is an opportunity

to arrive at a similar interpretation for ν' in terms of the slope of a linear best fit of the observed log-return data when the time dimension is transformed under this power law parametrised by the random variables t_c and β. This idea is explored in the next section.

7.1.2 Maximum log-likelihood analysis of the power-law hazard function

In Chapter 6, the observed log-price returns of an asset were given by $\mathbf{r} = (r_1, r_2, \ldots, r_n)$, where $r_1 = 0$, and $r_i = x_i - x_{i-1}$, for $i = 1, \ldots, n$, where $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ are the asset log-prices at times $t_1, t_2, \ldots, t_n$. Given the power law model, a simplified probability density function, $\tilde{F}(r, t)$, is derived as

$$\tilde{F}(r_i, t_i) = \frac{1}{\sqrt{2\pi\sigma^2\Delta_t}} \exp\left[-\frac{(r_i - \Delta_t(\mu - \nu(t_c - t_i)^\beta))^2}{2\sigma^2\Delta_t}\right], \tag{7.11}$$

where Δ_t is a scaling factor to take into account μ and ν being expressed as annualised values. Then, in determining the maximum log-likelihood it was found that the maximising value of ν, i.e., ν' can be given by

$$\nu' = -\frac{1}{\Delta_t} \frac{n\left(\sum_{i=1}^n (t_c - t_i)^\beta r_i - \bar{\mathbf{r}} \sum_{i=1}^n (t_c - t_i)^\beta\right)}{n\sum_{i=1}^n (t_c - t_i)^{2\beta} - \left(\sum_{i=1}^n (t_c - t_i)^\beta\right)^2}. \tag{7.12}$$

where $\bar{\mathbf{r}}$ is the sample mean of the set of observed log-returns, $\mathbf{r}$. Consider the change of variable

$$\hat{T}_i = (t_c - t_i)^\beta, \tag{7.13}$$

where $\hat{T}_i$ is a random variable. The equation (7.12) then becomes

$$\nu' = -\frac{1}{\Delta_t} \frac{n\left(\sum_{i=1}^n \hat{T}_i r_i - \bar{\mathbf{r}} \sum_{i=1}^n \hat{T}_i\right)}{n\sum_{i=1}^n \hat{T}_i^2 - \left(\sum_{i=1}^n \hat{T}_i\right)^2} \tag{7.14}$$

so that, as seen in the previous section,

$$\nu' = -\frac{1}{\Delta_t} \frac{\mathrm{Cov}(\hat{\mathbf{T}}, \mathbf{r})}{\mathrm{Var}(\hat{\mathbf{T}})}, \tag{7.15}$$

so that ν' can be interpreted as the log-likelihood maximised slope of the linear fit of the observed data against the observation time transformed under the map in equation (7.13). *This suggests that a bubble may be considered as a constant rate of return in what could be described as "bubble time".*

However, each $\hat{T}_i$ is a random variable, being a function of the random variables, β and t_c. Now, if one writes $\theta_i = t_c - t_i$ so that,

$$\hat{T}_i = \theta_i^{\beta} \tag{7.16}$$

then, recalling from §6.2.1 that $-1 < \beta < 0$ such that the asset has a finite value but infinite derivative at $t_i = t_c$, one may let θ_i and β be random variables such that

$$\beta \sim U(-1, 0), \quad \theta_i \sim U(s_0, s_1) \tag{7.17}$$

where s_0 is the shortest time period between the critical time, t_c, and the observed time t_i, and s_1 is the longest. In general, s_0 will be equal to the time between t_i and the trading period following the last observation date of the data set in question, and s_1 will be the value of s_0 plus a fixed length of time that is chosen to limit the critical time beyond the end of the observed data set.

Given that, with suitable assumptions for the values of s_0 and s_1, it is possible to determine the expected values of each $\hat{T}_i$ analytically. This presents an opportunity to consider a proxy for ν' that is derived in terms of $\mathrm{E}\left[\hat{T}\right]$. If the proxy measure is denoted $\bar{\nu}'$ then it is defined that

$$\bar{\nu}' \triangleq -\frac{1}{\Delta_t} \frac{n\left(\sum_{i=1}^{n} \mathrm{E}\left[\hat{T}_i\right] r_i - \bar{\mathbf{r}} \sum_{i=1}^{n} \mathrm{E}\left[\hat{T}_i\right]\right)}{n \sum_{i=1}^{n} \mathrm{E}\left[\hat{T}_i\right]^2 - \left(\sum_{i=1}^{n} \mathrm{E}\left[\hat{T}_i\right]\right)^2} \tag{7.18}$$

or, expressed more simply as a ratio of sample covariance and sample variance

$$\bar{\nu}' = -\frac{1}{\Delta_t} \frac{\mathrm{Cov}\left(\hat{\mathbf{t}}, \mathbf{r}\right)}{\mathrm{Var}\left(\hat{\mathbf{t}}\right)} \tag{7.19}$$

where $\hat{\mathbf{t}} = \left(\mathrm{E}\left[\hat{T}_1\right], \mathrm{E}\left[\hat{T}_2\right], \ldots, \mathrm{E}\left[\hat{T}_n\right]\right)$. Therefore, if it is possible to derive an

analytical expression for each $\mathrm{E}\left[\hat{T}_i\right]$, then calculation of the parameter $\bar{\nu}'$ becomes straightforward, and the comparative computational expense is greatly reduced.

7.2 Expectation of $\hat{T}_i$

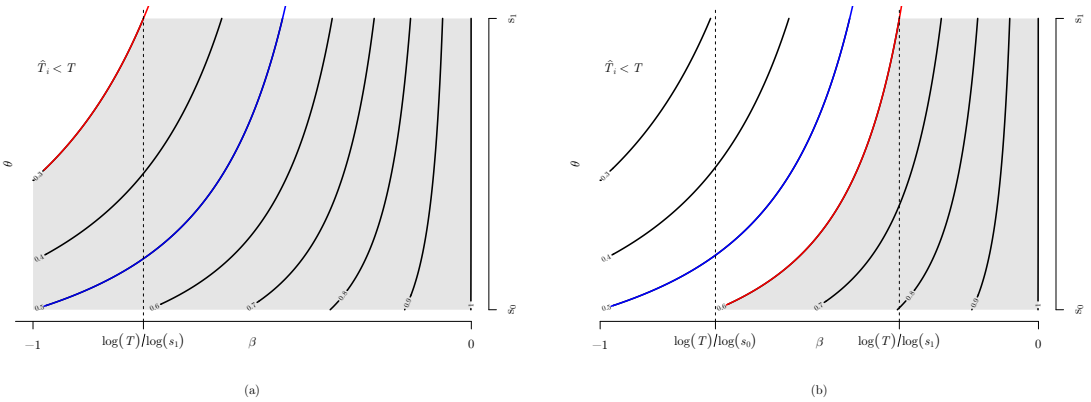
The aim is to derive an analytical expression for each expected value of the transformed time, $\mathrm{E}\left[\hat{T}_i\right]$, given the distributions of the random variables β and θ. Choosing an arbitrary value, T, and allowing $\hat{T}_i = T$ then θ can be written as a function of β as

$$\theta_i(\beta) = T^{\frac{1}{\beta}} \tag{7.20}$$

If s_0 is the shortest possible time (measured in years) between a particular instance of t_i and the critical time t_c, and s_1 is the longest, then there are various scenarios one need consider when determining the resulting distribution of the transformed time $\hat{T}_i$ and in turn, the expected value. These scenarios depend on the chosen values of s_0 and s_1, and in particular whether these values are both less than 1, both greater than 1, or span 1. Additionally, one must consider those scenarios where $s_0 = 1$ and where $s_1 = 1$. For each of these scenarios, contours of constant value of T are considered for the range of possible values of θ_i and β. The area bounded by these extrema where $\hat{T}_i \leq T$ defines the probability $P(\hat{T}_i \leq T)$ for the scenario in question. Once this probability distribution is determined, it may be possible to analytically determine the expected value of the distribution.

7.2.1 Scenario 1

In this scenario $s_1 > s_0 > 1$. Therefore the range of possible values of $\hat{T}_i$ are given by $1/s_1 \leq \hat{T}_i < 1$. In order to determine the probability $P(\hat{T}_i \leq T)$, one can find the complement of the area which is bounded by the curve $\theta_i = T^{1/\beta} - s_0$ and the line $\theta_i = s_1$ (see Fig. 7.1). In this scenario, there are two cases that must be considered to calculate the probability for all possible values of T. Firstly, the case where values of T are such that the curve $\theta_i = T^{1/\beta}$ intersects the line $\theta = s_0$ when $\beta \leq -1$, where it can be seen that for this condition to hold $1/s_1 \leq T \leq 1/s_0$,



(a) (b)

Figure 7.1: Scenario 1, where $s_1 > s_0 > 1$. These plots take $s_0 = 2$ and $s_1 = 5$ for illustrative purposes, and show contours for values of T where $\theta = T^{1/\beta}$. In order to determine the probability $P(\hat{T}_i < T)$ one must evaluate the area of the unshaded areas. It can be seen from this figure that there are two cases that need to be considered in order to evaluate this probability, and these case are separated by the value $\hat{T}_i = 1/s_0$, shown as the blue line in the above plots.

Fig. 7.1(a). Secondly, where values of T are such that the curve intersects the line $\theta = s_0$ when $\beta \geq -1$. In this case values of T are in the range $[1/s_0, 1]$, Fig. 7.1(b).

Case 1: $1/s_1 \leq T \leq 1/s_0$. In this case, with reference to Fig. 7.1(a), the probability, $P(\hat{T}_i \leq T)$, is given by

$$P(\hat{T}_i \leq T) = 1 - \frac{1}{\alpha}\left[\int_{-1}^{\phi_1(T)} (T^{\frac{1}{\beta}} - s_0)d\beta - \alpha\phi_1\right], \tag{7.21}$$

where $\alpha = s_1 - s_0$ and $\phi_1(T) = \log T/\log s_1$. Since the probability density function, $f(x)$, is the derivative of the cumulative distribution function, it follows the probability density function of $\hat{T}_i$ is therefore given by

$$\begin{aligned} f_{\hat{T}}(T) &= \frac{\partial}{\partial T}\left[1 + \phi_1 - \frac{1}{\alpha}\int_{-1}^{\phi_1(T)} (T^{\frac{1}{\beta}} - s_0)d\beta\right] \\ &= \frac{d\phi_1}{dT} - \frac{1}{\alpha}\frac{\partial}{\partial T}\int_{-1}^{\phi_1(T)} (T^{\frac{1}{\beta}} - s_0)d\beta\,. \end{aligned} \tag{7.22}$$

Using Leibniz's rule

$$\frac{\partial}{\partial x}\left(\int_{a(x)}^{b(x)} f(x,t)dt\right) = f(x,b(x))\frac{d}{dx}b(x) - f(x,a(x))\frac{d}{dx}a(x) + \int_{a(x)}^{b(x)}\frac{\partial}{\partial x}f(x,t)dt \tag{7.23}$$

to differentiate under the integral, and noting

$$\frac{d\phi_0}{dT} = \frac{1}{T\log s_0}, \quad \frac{d\phi_1}{dT} = \frac{1}{T\log s_1}\,, \tag{7.24}$$

one observes that

$$T^{\frac{1}{\phi_i}} = T^{\frac{\log s_i}{\log T}} = s_i, \quad i = 1,2 \tag{7.25}$$

Therefore, one may write

$$\begin{aligned}\frac{\partial}{\partial T}\int_{-1}^{\phi_1(T)}(T^{\frac{1}{\beta}} - s_0)d\beta &= \frac{s_1 - s_0}{T\log s_1} + \int_{-1}^{\phi_1(T)}\frac{\partial}{\partial T}(T^{\frac{1}{\beta}} - s_0)d\beta \\ &= \frac{\alpha}{T\log s_1} + \int_{-1}^{\phi_1(T)}\frac{T^{\frac{1}{\beta}}}{T\beta}d\beta\,.\end{aligned} \tag{7.26}$$

Therefore, this give the result

$$f_{\hat{T}}(T) = -\frac{1}{\alpha}\int_{-1}^{\phi_1(T)}\frac{T^{\frac{1}{\beta}}}{T\beta}d\beta\,. \tag{7.27}$$

Case 2: $1/s_0 \leq T \leq 1$. In this case the probability, $P(\hat{T}_i \leq T)$ is given by

$$P(\hat{T}_i \leq T) = 1 - \frac{1}{\alpha}\left[\int_{\phi_0(T)}^{\phi_1(T)}(T^{\frac{1}{\beta}} - s_0)d\beta - \alpha\phi_1\right] \tag{7.28}$$

where $\phi_0(T) = \log T/\log s_0$. It follows, as before, that the probability density function of $\hat{T}_i$ is given by

$$\begin{aligned}f_{\hat{T}_i}(T) &= \frac{\partial}{\partial T}\left[1 + \phi_1 - \frac{1}{\alpha}\int_{\phi_0(T)}^{\phi_1(T)}(T^{\frac{1}{\beta}} - s_0)d\beta\right] \\ &= \frac{d\phi_1}{dT} - \frac{1}{\alpha}\frac{\partial}{\partial T}\int_{\phi_0(T)}^{\phi_1(T)}(T^{\frac{1}{\beta}} - s_0)d\beta\end{aligned} \tag{7.29}$$

Again, differentiating under the integral gives as in the previous case gives, since $T^{1/\phi_0(T)} = s_0$,

$$\begin{aligned}\frac{\partial}{\partial T}\int_{\phi_0(T)}^{\phi_1(T)}(T^{\frac{1}{\beta}}-s_0)d\beta &= \frac{s_1-s_0}{T\log s_1}+\int_{\phi_0(T)}^{\phi_1(T)}\frac{\partial}{\partial T}(T^{\frac{1}{\beta}}-s_0)d\beta \\ &= \frac{\alpha}{T\log s_1}+\int_{\phi_0(T)}^{\phi_1(T)}\frac{T^{\frac{1}{\beta}}}{T\beta}d\beta \end{aligned} \tag{7.30}$$

so that

$$f_{\hat{T}_i}(T) = -\frac{1}{\alpha}\int_{\phi_0(T)}^{\phi_1(T)}\frac{T^{\frac{1}{\beta}}}{T\beta}d\beta\,. \tag{7.31}$$

Now, given the probability density function is piecewise in T, the expected value of T, $\mathrm{E}[T]$, can be written as

$$\begin{aligned}\mathrm{E}[T] &= \int_{\frac{1}{s_1}}^{1} T f_{\hat{T}_i}(T)dT \\ &= \int_{\frac{1}{s_1}}^{\frac{1}{s_0}} T f_{\hat{T}_i}(T)dT + \int_{\frac{1}{s_0}}^{1} T f_{\hat{T}_i}(T)dT \\ &= -\frac{1}{\alpha}\left[\int_{\frac{1}{s_1}}^{\frac{1}{s_0}}\int_{-1}^{\phi_1(T)}\frac{T^{\frac{1}{\beta}}}{\beta}d\beta dT + \int_{\frac{1}{s_0}}^{1}\int_{\phi_0(T)}^{\phi_1(T)}\frac{T^{\frac{1}{\beta}}}{\beta}d\beta dT\right]. \end{aligned} \tag{7.32}$$

Making the change of variable $\lambda = -\log T/\beta$ the expression is simplified to

$$\mathrm{E}[T] = -\frac{1}{\alpha}\left[\int_{\frac{1}{s_1}}^{\frac{1}{s_0}}\int_{\log T}^{-\log s_1}-\frac{e^{-\lambda}}{\lambda}d\lambda dT + \int_{\frac{1}{s_0}}^{1}\int_{-\log s_0}^{-\log s_1}-\frac{e^{-\lambda}}{\lambda}d\lambda dT\right]. \tag{7.33}$$

However, the definition of the exponential integral for non-zero values of x can be written as

$$\mathrm{Ei}[x] = -\int_{-x}^{\infty}\frac{e^{-t}}{t}dt \tag{7.34}$$

so that if a, b are positive real numbers where $a < b$, it can be seen that

$$\begin{aligned}-\int_{-b}^{-a}\frac{e^{-t}}{t}dt &= -\left(\int_{-b}^{\infty}\frac{e^{-t}}{t}dt - \int_{-a}^{\infty}\frac{e^{-t}}{t}dt\right) \\ &= \mathrm{Ei}[b]-\mathrm{Ei}[a]\,. \end{aligned} \tag{7.35}$$

Therefore, one may simplify the expectation expression to

$$\begin{aligned}\mathrm{E}[T] = \frac{1}{\alpha}\int_{\frac{1}{s_1}}^{\frac{1}{s_0}} \left(\mathrm{Ei}[\log s_1] - \mathrm{Ei}[-\log T]\right) dT \\ + \frac{1}{\alpha}\int_{\frac{1}{s_0}}^{1} \left(\mathrm{Ei}[\log s_1] - \mathrm{Ei}[\log s_0]\right) dT\,.\end{aligned} \tag{7.36}$$

Furthermore, since s_0 and s_1 are constants, the integrals in the expression are reduce by

$$\mathrm{E}[T] = \frac{1}{\alpha}\left(1 - \frac{1}{s_1}\right)\mathrm{Ei}[\log s_1] + \frac{1}{\alpha}\left(\frac{1}{s_0} - 1\right)\mathrm{Ei}[\log s_0] - \frac{1}{\alpha}\int_{\frac{1}{s_1}}^{\frac{1}{s_0}} \mathrm{Ei}[-\log T]dT\,. \tag{7.37}$$

Since

$$\frac{d}{dT}\,\mathrm{Ei}[-\log T] = \frac{1}{T^2 \log T}\,, \tag{7.38}$$

one may integrate the remaining integral by parts such that

$$\begin{aligned}\int_{\frac{1}{s_1}}^{\frac{1}{s_0}} \mathrm{Ei}[-\log T]dT &= \left[T\,\mathrm{Ei}[-\log T]\right]_{\frac{1}{s_1}}^{\frac{1}{s_0}} - \int_{\frac{1}{s_1}}^{\frac{1}{s_0}} \frac{dT}{T\log T} \\ &= \left[T\,\mathrm{Ei}[-\log T] - \log(\log T)\right]_{\frac{1}{s_1}}^{\frac{1}{s_0}} \\ &= \frac{\mathrm{Ei}[\log s_0]}{s_0} - \frac{\mathrm{Ei}[\log s_1]}{s_1} - \log\left(-\log s_0\right) + \log\left(-\log s_1\right)\end{aligned} \tag{7.39}$$

which, for the case where $s_1 > s_0 > 1$, gives the expected value

$$\mathrm{E}[T] = \frac{1}{\alpha}\left(\mathrm{Ei}[\log s_1] - \mathrm{Ei}[\log s_0] + \log\left(\frac{\log s_0}{\log s_1}\right)\right). \tag{7.40}$$

7.2.2 Scenario 2

In this scenario, $s_0 < 1$ and $s_1 > 1$. It can be see from Fig. 7.2 that $1/s_1 \leq T \leq 1/s_0$. As before, there are two cases to consider, and as can be seen in Fig. 7.2, these cases are separated by the line $T = 1$. When $T < 1$, the lower bound on T is given by $1/s_1$, and conversely, when $T > 1$, the upper bound is given by $1/s_0$.

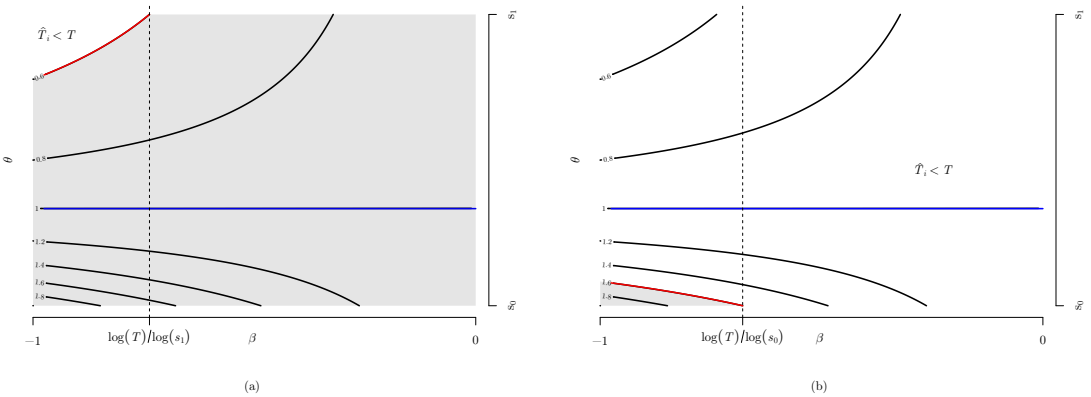


Figure 7.2: Scenario 2, where $s_0 < 1$, $s_1 > 1$. These plots take $s_0 = 0.5$ and $s_1 = 2$ for illustrative purposes, and show contours for values of T where $\theta = T^{1/\beta}$. As before, the probability $P(\hat{T}_i < T)$ is given by the unshaded areas, and clearly there are two cases that need to be considered. These case are separated by the value $\hat{T}_i = 1$, shown as the blue line in the above plots.

Case 1: $1/s_1 \leq T \leq 1$. With reference to Fig. 7.2(a) the probability, $P(\hat{T}_i \leq T)$ in this case is given by

$$P(\hat{T}_i \leq T) = 1 - \frac{1}{\alpha}\left[\int_{-1}^{\phi_1(T)} (T^{\frac{1}{\beta}} - s_0)d\beta - \alpha\phi_1\right] . \tag{7.41}$$

Case 2: $1 \leq T \leq 1/s_0$. The probability, $P(\hat{T}_i \leq T)$ in this case, Fig. 7.2(b), is very similar to that given in Case 1 of this scenario, and can simply be written as

$$P(\hat{T}_i \leq T) = 1 - \frac{1}{\alpha}\left[\int_{-1}^{\phi_0(T)} (T^{\frac{1}{\beta}} - s_0)d\beta\right] . \tag{7.42}$$

For Case 1, the probability density function is, as before, found by differentiating the cumulative probability function, i.e.

$$f_{\hat{T}_i}(T) = \frac{\partial}{\partial T}\left[1 - \frac{1}{\alpha}\left(\int_{-1}^{\phi_1(T)} (T^{\frac{1}{\beta}} - s_0)d\beta - \alpha\phi_1\right)\right] . \tag{7.43}$$

Since it has been seen that one may differentiate under the integral such that

$$\frac{\partial}{\partial T}\int_{-1}^{\phi_1(T)} (T^{\frac{1}{\beta}} - s_0)d\beta = \frac{\alpha}{T\log s_1} + \int_{-1}^{\phi_1(T)} \frac{T^{\frac{1}{\beta}}}{T\beta}d\beta\,, \tag{7.44}$$

it becomes clear that the probability density function in this Case 1 can be written as

$$f_{\hat{T}_i}(T) = -\frac{1}{\alpha}\int_{-1}^{\phi_1(T)} \frac{T^{\frac{1}{\beta}}}{T\beta} d\beta \,. \tag{7.45}$$

Since it is also apparent that

$$\frac{\partial}{\partial T}\int_{-1}^{\phi_0(T)} (T^{\frac{1}{\beta}} - s_0) d\beta = \int_{-1}^{\phi_0(T)} \frac{T^{\frac{1}{\beta}}}{T\beta} d\beta \,, \tag{7.46}$$

one may immediately write the probability density functions for Case 2 as

$$f_{\hat{T}_i}(T) = -\frac{1}{\alpha}\int_{-1}^{\phi_0(T)} \frac{T^{\frac{1}{\beta}}}{T\beta} d\beta \,. \tag{7.47}$$

Therefore, the population mean $\mathrm{E}[T]$ may expressed as

$$\mathrm{E}[T] = -\frac{1}{\alpha}\left[\int_{\frac{1}{s_1}}^{1}\int_{-1}^{\phi_1(T)} \frac{T^{\frac{1}{\beta}}}{\beta} d\beta dT + \int_{1}^{\frac{1}{s_0}}\int_{-1}^{\phi_0(T)} \frac{T^{\frac{1}{\beta}}}{\beta} d\beta dT\right] \tag{7.48}$$

which, by making the same change of variable as in (7.33), can be written in terms of the exponential integral as follows

$$\mathrm{E}[T] = -\frac{1}{\alpha}\left[\int_{\frac{1}{s_1}}^{\frac{1}{s_0}}\int_{\log T}^{-\log s_1} -\frac{e^{-\lambda}}{\lambda} d\lambda dT + \int_{\frac{1}{s_0}}^{1}\int_{\log T}^{-\log s_1} -\frac{e^{-\lambda}}{\lambda} d\lambda dT\right] . \tag{7.49}$$

Then combining with (7.35),

$$\begin{aligned}\mathrm{E}[T] = \frac{1}{\alpha}\int_{\frac{1}{s_1}}^{1} \left(\mathrm{Ei}[\log s_1] - \mathrm{Ei}[-\log T]\right) dT \\ + \frac{1}{\alpha}\int_{1}^{\frac{1}{s_0}} \left(\mathrm{Ei}[\log s_0] - \mathrm{Ei}[-\log T]\right) dT\end{aligned} \,. \tag{7.50}$$

Since s_0 and s_1 are constants, this simplifies to

$$\mathrm{E}[T] = \frac{1}{\alpha}\left(\mathrm{Ei}[\log s_1] - \mathrm{Ei}[\log s_0] - \int_{\frac{1}{s_1}}^{\frac{1}{s_0}} \mathrm{Ei}[-\log T] dT\right) . \tag{7.51}$$

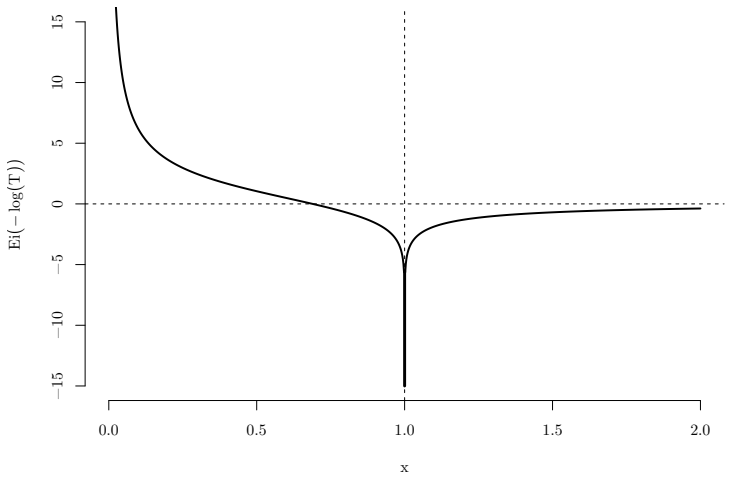


Figure 7.3: Plot of $\mathrm{Ei}[-\log x]$ near $x = 1$. The indefinite integral of $\mathrm{Ei}[-\log x]$ is given by $\mathrm{Ei}[-\log x]x - \log(\log(x))$. This is undefined on the real line for $x \leq 1$. However, it is possible to evaluate the definite integral across the singularity since $\lim_{\epsilon\to 0} \log\left(\frac{\log(x-\epsilon)}{\log(x+\epsilon)}\right) = 0$.

By integrating $\mathrm{Ei}[-\log T]$ across the singularity at $T = 1$ (Fig. 7.3), this expression simplifies to

$$\mathrm{E}[T] = \frac{1}{\alpha}\left(\mathrm{Ei}[\log s_1] - \mathrm{Ei}[\log s_0] + \log\left(-\frac{\log s_0}{\log s_1}\right)\right) . \tag{7.52}$$

7.2.3 Scenario 3

In this scenario, $s_0 < s_1 < 1$. From Fig. 7.4 it can be seen that $1 \leq \hat{T}_i \leq 1/s_0$. As before there are two cases to consider, that where $1 \leq T \leq 1/s_1$, Fig. 7.4(a), and the other where $1/s_1 \leq T \leq 1/s_0$, Fig. 7.4(b).

Case 1: $1/s_0 \leq T \leq 1/s_1$. In this first case where the cumulative probability function can be written

$$P(\hat{T}_i \leq T) = 1 - \frac{1}{\alpha}\left[\int_{-1}^{\phi_0(T)} (T^{\frac{1}{\beta}} - s_0)d\beta\right] . \tag{7.53}$$

This is exactly the same as that derived in scenario 2, Case 2. Therefore, the probability density function for this case may be written as

$$f_{\hat{T}_i}(T) = -\frac{1}{\alpha}\int_{-1}^{\phi_0(T)} \frac{T^{\frac{1}{\beta}}}{T\beta} d\beta \,. \tag{7.54}$$

Case 2: $1 \leq T \leq 1/s_1$. The cumulative probability function for the second case can be written as

$$\begin{aligned} P(\hat{T}_i \leq T) &= 1 - \frac{1}{\alpha}\left[\alpha(1+\phi_1) + \int_{\phi_1(T)}^{\phi_0(T)} (T^{\frac{1}{\beta}} - s_0) d\beta\right] \\ &= -\phi_1 - \frac{1}{\alpha}\int_{\phi_1(T)}^{\phi_0(T)} (T^{\frac{1}{\beta}} - s_0) d\beta \,. \end{aligned} \tag{7.55}$$

Therefore,

$$\begin{aligned} f_{\hat{T}_i}(T) &= -\frac{\partial}{\partial T}\left[\phi_1 + \frac{1}{\alpha}\int_{\phi_1(T)}^{\phi_0(T)} (T^{\frac{1}{\beta}} - s_0) d\beta\right] \\ &= -\frac{d\phi_1}{dT} - \frac{1}{\alpha}\frac{\partial}{\partial T}\int_{\phi_1(T)}^{\phi_0(T)} (T^{\frac{1}{\beta}} - s_0) d\beta \\ &= -\frac{1}{\alpha}\int_{\phi_1(T)}^{\phi_0(T)} \frac{T^{\frac{1}{\beta}}}{T\beta} d\beta \,. \end{aligned} \tag{7.56}$$

Therefore, the expected value of T, $\mathrm{E}[T]$, in this scenario is given by

$$\mathrm{E}[T] = -\frac{1}{\alpha}\left[\int_1^{\frac{1}{s_1}} \int_{\phi_1(T)}^{\phi_0(T)} \frac{T^{\frac{1}{\beta}}}{\beta} d\beta dT + \int_{\frac{1}{s_1}}^{\frac{1}{s_0}} \int_{-1}^{\phi_0(T)} \frac{T^{\frac{1}{\beta}}}{\beta} d\beta dT\right] . \tag{7.57}$$

Following the same change of variable as in the previous two scenarios, it can be shown that

$$\begin{aligned} \mathrm{E}[T] = \frac{1}{\alpha}\int_1^{\frac{1}{s_1}} &\left(\mathrm{Ei}[\log s_0] - \mathrm{Ei}[\log s_1]\right) dT \\ &+ \frac{1}{\alpha}\int_{\frac{1}{s_1}}^{\frac{1}{s_0}} \left(\mathrm{Ei}[\log s_0] - \mathrm{Ei}[-\log T]\right) dT \end{aligned} \,. \tag{7.58}$$

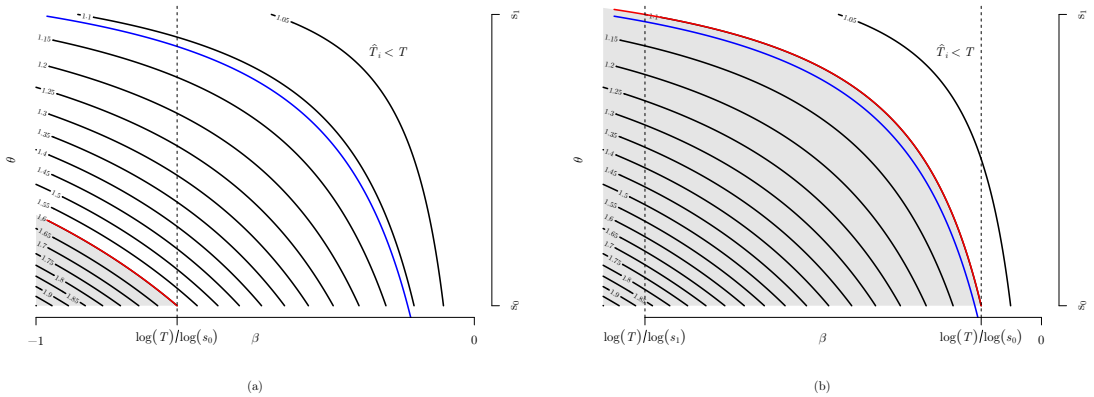


Figure 7.4: Scenario 3, $s_0 < s_1 < 1$. These plots take $s_0 = 0.5$ and $s_1 = 0.9$ for illustrative purposes, and show contours for values of T where $\theta = T^{1/\beta}$. Again, the probability $P(\hat{T}_i < T)$ is given by the unshaded areas, and two cases are considered. These cases are separated by the value $\hat{T}_i = 1$, shown as the blue line in the above plots.

Therefore,

$$\mathrm{E}[T] = \frac{1}{\alpha}\left[\mathrm{Ei}[\log s_1]\left(1+\frac{1}{s_1}\right) - \mathrm{Ei}[\log s_0]\left(1+\frac{1}{s_0}\right) - \int_{\frac{1}{s_1}}^{\frac{1}{s_0}} \mathrm{Ei}[-\log T]dT\right] \tag{7.59}$$

and from (7.39) it can be seen that

$$\mathrm{E}[T] = \frac{1}{\alpha}\left(\mathrm{Ei}[\log s_1] - \mathrm{Ei}[\log s_0] + \log\left(\frac{\log s_0}{\log s_1}\right)\right). \tag{7.60}$$

7.2.4 Scenario 4

In this scenario there are two case, one where $s_0 = 1$ and $s_1 > 1$ and the other where $s_1 = 1$ and $s_0 < 1$. It can see from Fig. 7.5 that in both these cases $1 \leq \hat{T}_i \leq 1/s_0$, where $1 \leq T \leq 1/s_0$ for $s_0 = 1$, and $1/s_1 \leq T \leq 1$ where $S_1 = 1$.

Case 1: $s_1 = 1$ and $1 \leq T \leq 1/s_0$. In this case the probability, $P(\hat{T}_i \leq T)$ is given by

$$P(\hat{T}_i \leq T) = 1 - \frac{1}{\alpha}\left[\int_{-1}^{\phi_0(T)} (T^{\frac{1}{\beta}} - s_0)d\beta\right]. \tag{7.61}$$

Case 2: $s_0 = 1$ and $1/s_1 \leq T \leq 1$. In this case the probability, $P(\hat{T}_i \leq T)$ is given by

$$P(\hat{T}_i \leq T) = 1 - \frac{1}{\alpha}\left[\int_{-1}^{\phi_1(T)} (T^{\frac{1}{\beta}} - s_0)d\beta - \alpha\phi_1\right] \tag{7.62}$$

so that, by following the same arguments as before the respective probability density function are given by

$$f_{\hat{T}_i}(T) = -\frac{1}{\alpha}\int_{-1}^{\phi_0(T)} \frac{T^{\frac{1}{\beta}}}{T\beta}d\beta, \quad f_{\hat{t}}(T) = -\frac{1}{\alpha}\int_{-1}^{\phi_1(T)} \frac{T^{\frac{1}{\beta}}}{T\beta}d\beta\,. \tag{7.63}$$

When $s_1 = 1$, it follows that $s_0 < 1$ and the mean $\mathrm{E}[T]$ is expressed as

$$\begin{aligned}
\mathrm{E}[T] &= -\frac{1}{\alpha}\left[\int_1^{\frac{1}{s_0}}\int_{-1}^{\phi_0(T)} \frac{T^{\frac{1}{\beta}}}{\beta}d\beta dT\right] \\
&= \frac{1}{\alpha}\int_1^{\frac{1}{s_0}} \left(\mathrm{Ei}[\log s_0] - \mathrm{Ei}[-\log T]\right) dT \\
&= \frac{1}{\alpha}\left(\frac{1}{s_0} - 1\right)\mathrm{Ei}[\log s_0] - \frac{1}{\alpha}\int_1^{\frac{1}{s_0}} \mathrm{Ei}[-\log T]dT\,.
\end{aligned} \tag{7.64}$$

Evaluating the integral in the above expression by parts, with the lower limit being ϵ rather than 1, gives

$$\int_\epsilon^{\frac{1}{s_0}} \mathrm{Ei}[-\log T]dT = \frac{1}{s_0}\mathrm{Ei}[\log s_0] - \epsilon\,\mathrm{Ei}[-\log\epsilon] - \log(\log s_0) + \log(\log\epsilon) \tag{7.65}$$

such that

$$\lim_{\epsilon\to 1}\left(\epsilon\,\mathrm{Ei}[-\log\epsilon] - \log(\log\epsilon)\right) = -\lim_{\zeta\to 0}\left(\log(\zeta) - \mathrm{Ei}[\zeta]\right). \tag{7.66}$$

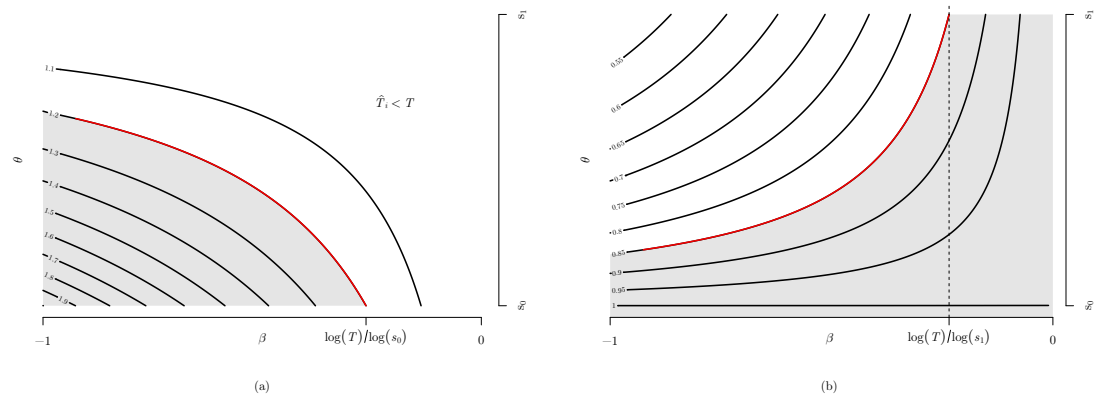


Figure 7.5: Scenario 4, left diagram $s_1 = 1, s_0 < 1$, and on the right $s_0 = 1, s_1 > 1$. For illustrative purposes, these plots take $s_0 = 0.5$ for the diagram on the left, and $s_1 = 2$ for the diagram on the right. Each plot shows contours for values of T where $\theta = T^{1/\beta}$. As in the previous plots, the probability $P(T_i \hat{<} T)$ is given by the unshaded areas.

Since $E_1[x] = -\operatorname{Ei}[x]$, it is clear from Gellar & Ng (1969), that this expression resolves to the Euler-Mascheroni constant, γ. Therefore

$$\mathrm{E}[T] = \frac{1}{\alpha}\left(\log(-\log s_0) - \operatorname{Ei}[\log s_0] - \gamma\right) . \tag{7.67}$$

Similarly, for $s_0 = 1$, the mean can be found to be

$$\mathrm{E}[T] = \frac{1}{\alpha}\left(\operatorname{Ei}[\log s_1] - \log(\log s_1) + \gamma\right) . \tag{7.68}$$

7.2.5 Piecewise expected value of $\hat{T}_i$

One may combine these scenarios in following the theorem:

Theorem 7.2.1. *If a random variable T is a function of two other random variables, β and θ, such that*

$$T = \theta_i^{\beta} \tag{7.69}$$

where β and θ are drawn from the distributions

$$\beta \sim U(-1, 0), \quad \theta_i \sim U(s_0, s_1) \tag{7.70}$$

where $s_0 < s_1$*, then*

$$\mathrm{E}[T] = \frac{1}{\alpha} \begin{cases} \mathrm{Ei}[\log s_1] - \log(\log s_1) + \gamma & s_0 = 1 < s_1 \\ \log(-\log s_0) - \mathrm{Ei}[\log s_0] - \gamma & s_0 < 1 = s_1 \\ \mathrm{Ei}[\log s_1] - \mathrm{Ei}[\log s_0] + \log\left(\frac{\log s_0}{\log s_1}\right) & s_0 < s_1 < 1 \\ \mathrm{Ei}[\log s_1] - \mathrm{Ei}[\log s_0] + \log\left(-\frac{\log s_0}{\log s_1}\right) & s_0 < 1 < s_1 \\ \mathrm{Ei}[\log s_1] - \mathrm{Ei}[\log s_0] + \log\left(\frac{\log s_0}{\log s_1}\right) & 1 < s_0 < s_1 \end{cases} \tag{7.71}$$

where $\alpha = s_1 - s_0$*.*

7.3 Application of $\bar{\nu}'$ in time series forecasting

Consider a time series where the time vector is given by $\mathbf{t} = (t_1, t_2, \ldots, t_n)$ and the observed data is given by $\mathbf{r} = (r_1, r_2, \ldots, r_n)$. In order to find values for $\bar{\nu}'$ at each data point, one first must decide how much historical data should be used to compute this value. In practice, this is very much a matter of experimentation, since exposing the calculation to too little data may not capture the development a superexponential growth regime, whereas too much data may indeed contaminate such a regime with periods of normal growth. However, for now, the number of data points, whatever they might be (up to to and including the last observation), is expressed as ω (not to be confused with the angular log-frequency ω used in the LPPLs). Therefore, from (7.15), for each data point the value of $\bar{\nu}'$ is given by

$$\bar{\nu}'_i = -\frac{1}{\Delta_t} \frac{\mathrm{Cov}(\hat{\mathbf{t}}_{i-\omega,i}, \mathbf{r}_{i-\omega,i})}{\mathrm{Var}(\hat{\mathbf{t}}_{i-\omega,i})}, \quad i > \omega \tag{7.72}$$

where $\hat{\mathbf{t}}_{i-\omega,i}$ is the vector $\left(\mathrm{E}\left[\hat{T}_{i-\omega}\right] \ldots \mathrm{E}\left[\hat{T}_i\right]\right)$ of expectations of observation times transformed by equation (7.13) and $\mathbf{r}_{i-\omega,i}$ is the vector of log-return observations $(r_{i-\omega} \ldots r_i)$.

For the purposes of this thesis, a conjecture is made such that as the values of $\bar{\nu}'$ become more negative, the market becomes more decisively governed by a power-law that typifies the development of market bubbles. This working hypoth-

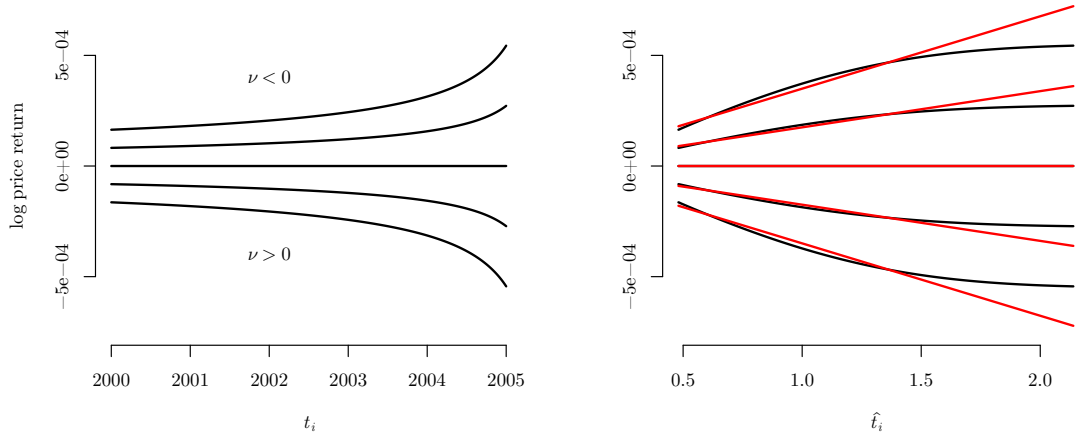


Figure 7.6: Synthetic time series plots with time transforms. The plot on the left shows synthetic log-price returns of a five-year period from 2000 to 2005 of a fictitious market that exhibits no stochastic features. One can see that if one sets ν to be negative the market shows superexponential growth, and when it is positive the market exhibits superexponential decline. On the right, the black line shows the market return plotted with the time axis transformed under (7.71). A linear fit is calculated for each synthetic time series and shown as the red lines. The slope of these linear fits is $-\bar{\nu}'$. The greater the slope, the greater the negative value of $\bar{\nu}'$.

esis is the cornerstone of the algorithm that is developed below. This important relationship is due to $\bar{\nu}'$ being the negative of the log-likelihood maximised slope of the linear fit of the market log-price returns versus the expected transformed time, i.e. the greater this slope, the greater the implied superexponential growth.

Similarly, as the values of $\bar{\nu}'$ become more positive, the working hypothesis is that this indicates the market is moving into a period of anti-bubble, again governed by a power law, but in this case with downward momentum. It should be noted here that the type of anti-bubble mentioned here is not analogous to anti-bubbles that have been discussed in the literature or, indeed, previously in Chapter 5 of this thesis. Historically, anti-bubbles have been thought of as regimes governed by power-laws of the form $(t - t_c)^\beta$, where the critical time t_c occurs *before* the onset of the market drawdown. This revised description of a power-law hazard function anti-bubble is that it is simply the opposite of a bubble. If, during a bubble regime, markets are driven by the probability of an expected drawdown of κ, then during an anti-bubble, downward motion is driven

by the probability of a market rally of κ after some future critical time.

It can be seen from the synthetic plots of Fig. 7.6 that transforming the log-return plots from normal time to time transformed by (7.71) does not result in linear forms, and that the measure $\bar{\nu}'$ is only an approximation of the "real" value of ν. Furthermore, when examining the root-mean-squared error of these approximations for varying the chosen values of ν, t_c and β it is apparent that these errors vary with parameter choices. As can be seen in Fig. 7.7 smaller r.m.s. errors occur in those circumstances where the actual value of t_c is further in the future, β is closer to zero, and the real value of ν is smaller. Conversely, those areas where t_c is closer to the end of the observed data, β is closer to minus 1 and the real value of ν is more negative lead to the greatest r.m.s. errors. The bias in $\bar{\nu}'$ as an estimator is accepted as a systemic deficiency in the underlying model, and it put to one side whist the predictive qualities of $\bar{\nu}'$ are determined.

7.3.1 Making a prediction

The working hypothesis of this thesis is that during a bubble regime, returns are driven higher due to participants requiring a better return for what is assumed to be increasing risk, i.e. as a bubble develops it becomes increasingly likely that the bubble will burst resulting in a crash with the market suffering a drawndown of κ. As described above, the converse of this is also held to be true in that during a period of anti-bubble behaviour, asset prices are driven further down as the predominantly "bearish" participants require increasing negative returns given the probability that the market will experience a significant "bounce" in the future.

The question, therefore, is whether the measure, $\bar{\nu}'_i$ is a good predictor of the direction of the market over any particular future investment horizon of τ data points. However, in order to make use of the value of $\bar{\nu}'$, it is necessary to condition this measure so that a comparison can be made as to what extent a value of $\bar{\nu}'_i$ should be considered as indicating a market positive or negative movement in the context of previously measured values. Therefore, for each data point, a moving window consisting of ω' previous observations of $\bar{\nu}'$ are taken and

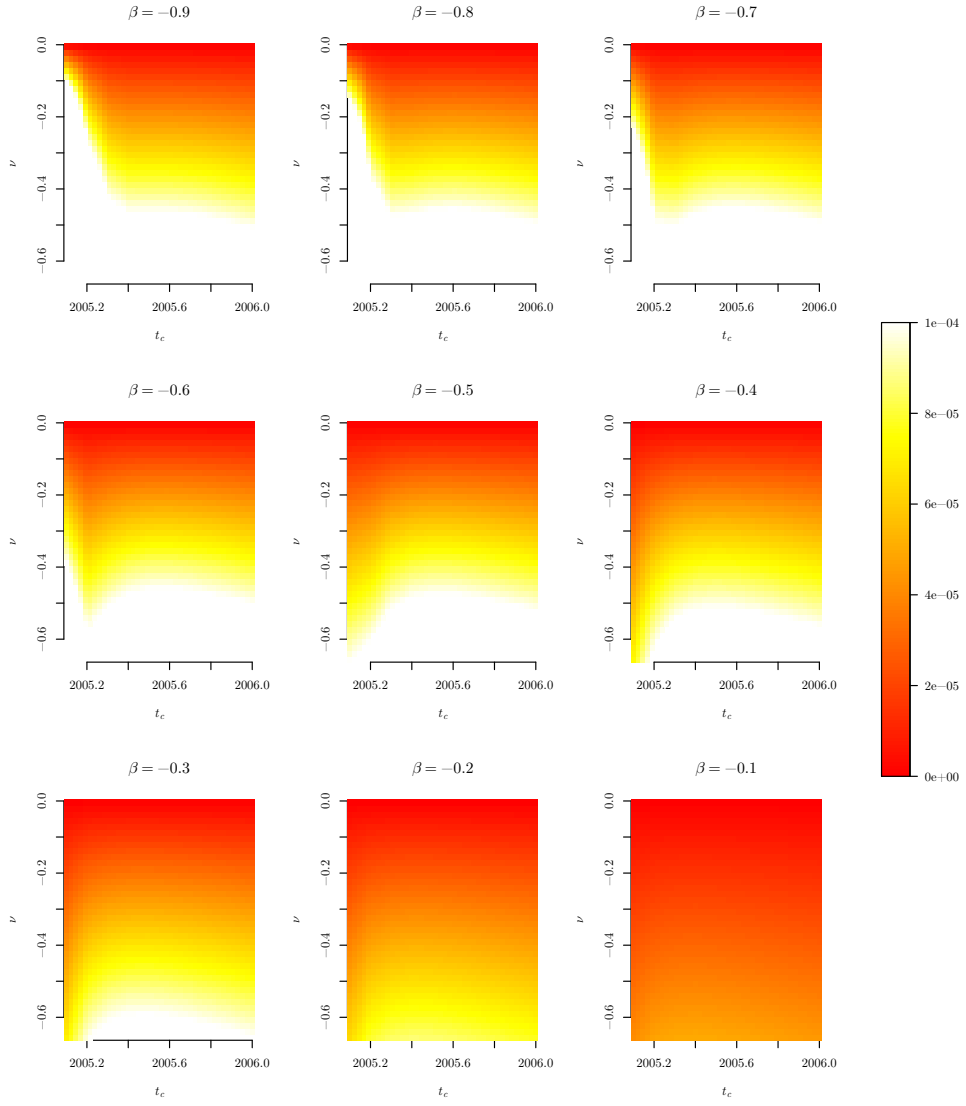


Figure 7.7: Root-mean-squared errors for five-year synthetic time series. By taking non-stochastic, synthetic time series as before, it can be seen that the root-mean-squared error (and hence the estimator bias and error variance) are dependant on chosen values of t_c, β and ν for each simulation. The plots show a heat-map of r.m.s. errors for varying t_c and ν, with each individual plot having a particular value of β. Areas in the red part of the spectrum have the smallest errors, whereas whiter shades indicate increasing errors. It can be seen that the smallest errors occur when t_c is further away from the last observed data point, β is closer to zero, and the chosen value of ν is relatively small compared to its possible values.

the empirical distribution function is determined by

$$\hat{F}_{\omega'}(\bar{\nu}'_i) = \frac{1}{\omega'} \sum_{j=i-\omega'}^{i} \mathbf{1}_{\bar{\nu}'_j \leq \bar{\nu}'_i} \tag{7.73}$$

where $\mathbf{1}_{\bar{\nu}'_j \leq \bar{\nu}'_i}$ is the indicator function defined as

$$\mathbf{1}_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases} \tag{7.74}$$

Under this construct, the lower the calculated value of $\hat{F}_{\omega'}(\bar{\nu}'_i)$, the more likely the market is in a bubble regime, and the more likely it is that the market will continue its growth. For higher calculated values the more likely the market is in an anti-bubble regime and the expectation is that the market will move lower in over some future investment horizon. Given this methodology, one can study to what extent the normalised $\bar{\nu}'_i$ parameter can predict the future direction of the market.

With this in mind, a methodology has been arrived at which aims to predict the direction of market movement over a pre-defined period. This methodology has similarities to modern machine learning algorithms in that, as time goes on, historical data is used to update parameters such that the algorithm "learns" from past experience. The steps of this methodology is as follows:

1. **Log return.** A time series of market data is taken, and for each data point, the difference between the log of the closing price at this data point and the log price at the previous point is taken. This is denoted r_i where $i = 1$ is the first data point in the time series.

2. **Convert to market time.** Each date in the time series is converted to "market time". Market time is the decimal expression of calendar dates where the number of trading days in each year is spread equally over that particular year.

3. **Select historical data period.** As described above, one must make a

decision about how much historical data is used to calculate the value of $\bar{\nu}'_i$. This value is arbitrarily set at $\omega = 1300$. With there being around 260 trading days in a year, this amounts to 5 years' worth of historical data for each $\bar{\nu}'_i$ calculation.

4. **Select historical normalisation period.** One also must decide over what historical periods the values of $\bar{\nu}'_i$ should be normalised. This is a rolling window of "market memory", i.e. the period where one assumes there is some sort of causal link between one market movement and the next. This is shorter than the period ω and arbitrarily chosen as six months, i.e. $\omega' = 130$.

5. **Calculate $\bar{\nu}'_i$ for each ith data point.** For each ith data point:

   (a) take the returns data $\mathbf{r} = (r_{i-\omega}, r_i)$;

   (b) transform the market time dates $\mathbf{t} = [t_{i-\omega}, t_i]$ using (7.71) with $s_0 = 0$ and $s_1 = 0.5$ (i.e. the critical time t_c is assumed to be within 6 months of the ith data point) to obtain the vector of time transformed expectations, $\hat{\mathbf{t}}_{i-\omega,i}$;

   (c) calculate $\bar{\nu}'_i = -\text{Cov}(\hat{\mathbf{t}}_i, \hat{r_i})/\text{Var}(\hat{\mathbf{t}}_i)$

6. **Normalise $\bar{\nu}'_i$.** Again, for each ith data point calculate the empirical distribution $\hat{F}_{\omega'}(\bar{\nu}'_i)$ given the previous ω' calculations of $\bar{\nu}'$. Denote each of these values as p_i.

7. **Point prediction.** Based on the *historical* performance of p_i as a predictor of market movement, a threshold level of this parameter is derived which is used to indicate whether the market price is predicted to be higher or lower τ trading periods after the ith observation. This method requires more in-depth study, and is described in detail in the following section.

7.3.2 Point by point prediction of market movement

At this point in the execution of the methodology, there exists a probability p_i for each ith observation which has been derived from the preceding ω' observations

of the $\bar{\nu}'$ values. The aim is to use p_i to predict whether at $t_{i+\tau}$ the market price will be higher or lower than the observation at time t_i. To do this one must decide at what value of p_i should the prediction switch from lower to higher. To determine the value of this threshold, one can examine how well the dataset of observations of p calculated in the interval $[i-\omega'-\tau, i-\tau]$ have predicted the set of binary responses

$$x_i = \begin{cases} 1 & y_i > y_{i-\tau} \\ 0 & y_i \leq y_{i-\tau} \end{cases} \tag{7.75}$$

where y_i is the closing market price at observation time t_i.

One can gauge the predictive qualities of a set of predictors in describing a set of binary responses by studying the receiver operating characteristic (ROC) curve. These graphs are routinely used in the the fields of machine learning, medical decision analysis and data science generally, and constitute a "technique for visualizing, organising and selecting classifiers" (Fawcett 2006).

If one considers an example of the sequence of binary responses, $\mathbf{x}$, i.e. actual observed data, calculated by (7.75) given the data set of market prices $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ and the set of predictors $\mathbf{p} = (p_1, p_2, \ldots, p_n)$, then applying some threshold value, $p_{\text{th}} \in [0, 1]$, to the set $\mathbf{p}$ such that

$$\hat{p}_i = \begin{cases} 1 & p_i > p_t \\ 0 & p_i \leq p_t \end{cases} \tag{7.76}$$

one can construct a so called confusion matrix (Table 7.1) mapping the actual values of the observations to the predicted values of the observations given the chosen value of p_{th}. For the purposes of this thesis, the actual outcomes are classified as one of the set (u', d'), denoting the fact that the market has moved either up or down over the chosen investment horizon, and the predicted outcome classifications are denoted by one of the set (u, d). In this way one is able to tabulate those bull market outcomes which have been successfully predicted by the classified response, a true positive (TP), and those bear market outcomes that have also been predicted correctly, a true negative (TN). Additionally, one

		Prediction outcome		
		u	**d**	**Total**
Actual value	**u′**	True Positive	False Negative	U′
	d′	False Positive	True Negative	D′
	Total	U	D	

Table 7.1: Confusion matrix of actual values and predicted outcomes. Actual, or observed, values are denoted by u' and d' for the market having moved up or down respectively from it current value over the investment period, τ. Predicted outcomes of up or down are denoted by u and d respectively. The number of times $u' = u$ gives the number of "True Positives" (TP). Similarly the number of "True Negatives" (TN), "False Positives" (FP) and "False Negatives" (FN) are given by $d' = d$, $d' = u$, and $u' = d$ respectively. It can be seen that total number of actual positive market moves is given by $U' = TP + FN$, and the total number of negative market moves by $D' = FP+TN$. Similarly, the total number of up predictions is given by $U = FP+FN$ and the number of predicted down by $D = FN + TN$.

is able to see the number of bull market prediction which turned out to be false, a false positive (FP), and those bear market predictions that incorrectly predicted a rising market, a false negative (FN). These totals allow one to derive a few measures that are of interest:

- Accuracy: $ACC = \frac{TP+TN}{U+D}$
- True positive rate, or sensitivity: $TPR = \frac{TP}{TP+FN}$
- False positive rate: $FPR = \frac{FP}{FP+TN}$
- True negative rate or specificity: $TNR = \frac{TN}{FP+TN} = 1 - FPR$

A ROC curve is simply 2-dimensional representation of the relationship between the sensitivity, usually plotted on the y-axis, and 1−specificity (i.e. the FPR), which is usually plotted on the x-axis, and can be seen a a graphical method of visualising the trade off between true positive and false positive predictions. For example, in the case at hand, one must decide whether the set $\mathbf{p}$ has predicted the response $\mathbf{y}$ is a statistically robust manner. If one decides that the

threshold at which $\mathbf{p}$ is classified to either a prediction of u or d is p_{th}, then one may plot the point (FPR, TPR) given the outcomes of the predictions for this value of the threshold. Clearly, for a prediction to be useful, it is desirable that this point should be above the line $y = x$, i.e. the true positive rate is greater than the false positive rate. However, if this point lies along the line $y = x$, this prediction is no better than chance for a binary response. Given this information about the predictive qualities of the predictor when the threshold value is set at the discrete value p_{th}, one may plot a continuous curve of all values of $p_{\text{th}} \in [0, 1]$, thereby showing how the predictor performs for all threshold values. An example of this methodology is shown for a particular set of data points in the S&P500 index time series in Fig. 7.8.

From this ROC curve one can see the most important factor from the perspective of the prediction algorithm, that of the optimal threshold. This is the threshold that produces the point with the greatest perpendicular distance from the main diagonal. This is where there is the best trade off between true and false positive results.

Additionally, it is also useful to consider the area under the ROC curve (AUC). This is a measure of how well the predictor is able to distinguish between the two binary classifications of actual outcomes. When the AUC is equal to unity, this is the ideal scenario, and the predictor can perfectly separate these two classes, whereas when the AUC is 0.5, the predictor has no ability to make any distinction, i.e. the predictor is no better than a random guess for all threshold values. When the AUC lies between 0.5 and 1, the AUC tells the user how well the model performs in predicting true positives and true negatives. If the AUC is less than 0.5, then the model is a reverse indicator. The AUC is not used in the algorithm as has been described above but, as will be seen later in this chapter, this measure is used to filter out predictions that are not based on firm statistical footings.

7.3.3 Example predictions - S&P500 index

In this section, an example of the practical use of the algorithm is presented along with its results. The example shows how one may use the algorithm to

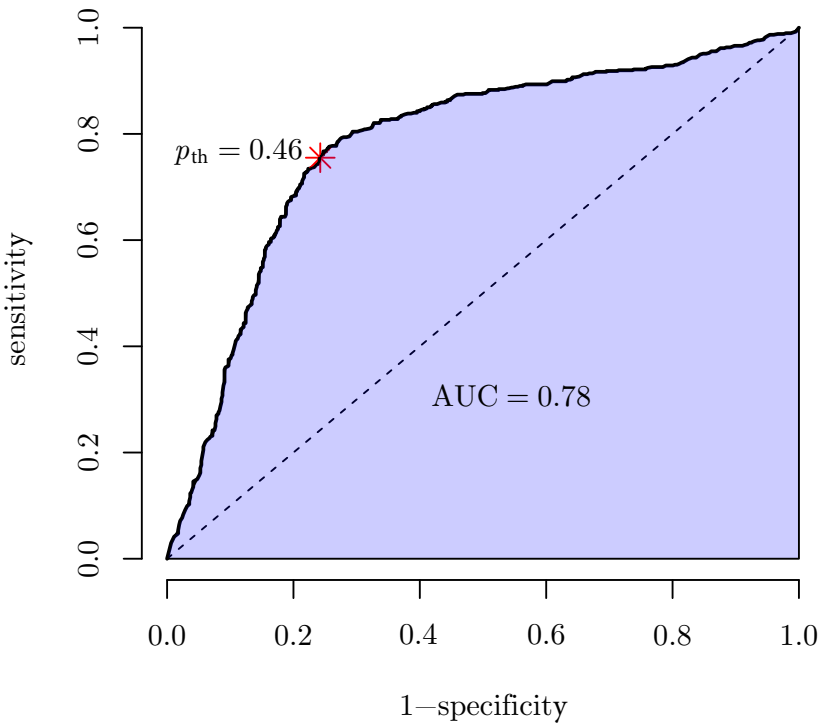


Figure 7.8: ROC curve for predictions for the S&P500 index for the five years from July 2001. Threshold values $p_{\text{th}} \in [0, 1]$ are used to map each p_i to a classification in (u, p) from which sensitivities and specificities are calculated and plotted on the graph. This results in a ROC curve from which an optimal value of p_{th} can be determined. Additionally the area under the ROC curve (AUC) can be calculated to give a measure of how effective the set $\mathbf{p}$ is in forecasting the response set $\mathbf{y}$. In this case the AUC is 0.78, which would indicate that the prediction is performing much better than chance, for which one would see a value of closer to 0.5 for the AUC. The optimum threshold value is $p_{\text{th}} = 0.46$ in this case. This means that the threshold that results in the point on the curve with the greatest perpendicular distance from the line $TP = FN$ is 0.46.

generate rapidly a historical set of back-tested performance statistics, and how the algorithm may be implemented to produce real-time predictions. It should be noted that there are a few parameters that must be chosen before the algorithm may be used, and the choice of these parameters can greatly affect the outcome. It is certainly true that one could optimise, or fine tune, these parameters so that more favourable back-tested results are achieved for each market investigated. However, in devising this methodology, a fixed set of parameters has been taken and used in all circumstances so that it cannot be argued that the most favourable were conditions were chosen for each experiment.

For this example, daily trading data from the S&P500 index was taken from online open sources (uk.finance.yahoo.com) going back to the early 1950's up until September 2019. In order to apply the methodology described in §7.3.1 and §7.3.1, the following (arbitrary) parameter choices were made:

- Historical data set length in data points: $\omega = 1300$
- Normalisation period in data points: $\omega' = 560$
- Minimum period to critical time in years: $s_0 = 0$
- Maximum period to critical time in years: $s_1 = 0.5$
- Investment horizon in data points: $\tau = 130$

Given these parameter choices, the algorithm was applied to a subset of the data starting at the beginning of 1995 and ending in September 2019 (a time period chosen such that results can be compared to other asset classes given the scope of their available data). Once the historical optimal threshold had been established for each trading date,this threshold was applied to the value of p_i calculated for that date to make a classification from the set (u, d), and a prediction was made of as to whether the market would be higher or lower after a period of τ years. The confusion matrix of the results are shown in Table 7.2.

As can be seen from these results, the methodology performs relatively well over this time period. The accuracy of the predictor is around 61% which would, by any measure in the financial markets sector, be a remarkable achievement for a

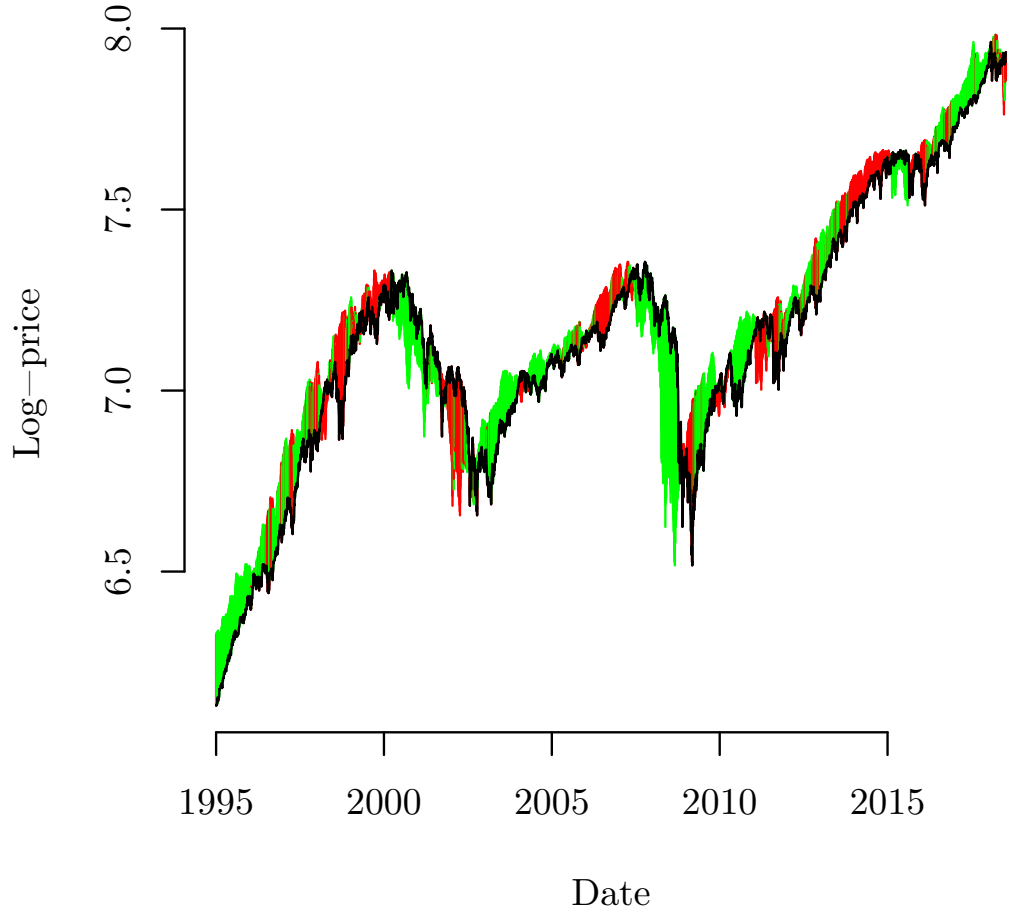


Figure 7.9: Back-testing results for 25 years of historical S&P500 index data. The algorithm is applied to 5 years worth of historical data for each data point, and the measure $\bar{\nu}'$ is normalised over a 2 year period. Having determined the historical optimum threshold value, p_{th} for each data point a forecast is made of whether the market will be higher or lower in at the end of the following 6 month period. For each data point a line extends either above or below the current market price (black line) to the value of the market price 6 months forward. If the line is green, the forecast for this data point has been successful, if red, the forecast is incorrect.

		Prediction outcome		
		Higher	**Lower**	**Total**
Actual value	**Higher**	2,624 0.605	1,712 0.395	4,336
	Lower	600 0.378	987 0.622	1,587
	Total	3,224	2,699	

Table 7.2: Confusion matrix for S&P500 index example. In this example, it can be seen that there were a total of $5,923$ observations, of which $4,336$ resulted in a higher asset price after the investment period, and $1,587$ were lower. Each element of the confusion matrix shows the absolute number of prediction outcomes, and underneath these numbers are expressed as percentages of each actual outcome. For instance, $TPR = 0.605$ and $TNR = 0.622$.

predictive model, and would certainly attract attention from proprietary trading desks. It can be also be seen from the results that the algorithm performs equally as well respect of both of the true positive and true negative rates.

However, the algorithm is designed only to make binary predictions of whether the market should rally or trade down over a certain period, and does not give any indication as to the expected size of market movement. Therefore, one cannot say that this algorithm is able to generate significant returns despite its high accuracy. One can gauge the profitability of the methodology by devising a very simple trading strategy and applying it to the predictions that have been made historically.

One way to do this is to imagine that once a prediction is made, a trader is able to open a position that has positive exposure to rising prices, i.e. go long the market, or to open a position that has positive exposure to falling prices, i.e. go short the market. This is indeed possible in many markets by entering into so called contracts for difference (**CFD**). By entering into CFD transactions the trader is making a "bet" with the exchange such that the trader will "win" or "lose" a certain amount based on how much the market moves from the time the position is opened. It is perfectly reasonable that a trader could bet, say, \$1 per

point that the S&P500 index rises, i.e. for every index point the index increases, the trader will make \$1. Conversely, if the market falls the trader will lose \$1 per point. Taking a "leveraged" position like this is particularly dangerous as one can lose more than the one's funds are sufficient to bear. If one buys an asset, the worst that can happen is the asset becomes worthless, but trading CFDs is very different in that as one loses money the exchange will ask for "margin", or cash collateral, to be deposited against these losses. If the trader is unable to meet the margin call the position is closed, and the trader will not be able to stay in the market to recover his/her losses.

Therefore, a simple strategy is constructed that allows a fictitious trader to apply the algorithm to a market of choice, and given the calculated threshold and value of p_i, to decide to go long or short the market for 1 currency unit per unit of price movement in the underlying CFD asset. It is assumed that there is no bid/offer spread, and that there are no trading costs associated with opening this position. If one back-tests this strategy over the last 25 years of the S&P500 index, one achieves the results shown in Fig. 7.11.

One can see from Fig. 7.9 that the strategy's predictive qualities are equally as effective in both bull and bear market periods. Another interesting feature of the predictions is that successful forecasts are very much clustered together, rather than being distributed throughout the time series. It is true that market prices trend in one direction or the other for long periods of time, so maybe this clustering in not wholly surprising, but the investment horizon that has been chosen (in this example, 6 months) is much shorter than the prediction clustering periods that are shown in this example, and, as such, one might expect that this smoothing would have effect over this time period rather then the longer clustering periods shown in the example.

Fig. 7.10 shows the distribution of gains and losses for each trade. As can be seen, in some circumstances, the fictitious trader can make very large gains with a single 1 currency unit per point bet, but equally, huge losses are also possible. In this example the median gain for a single trade is 48 currency units. However, as a practical trading strategy, this is only suitable for those fictitious traders

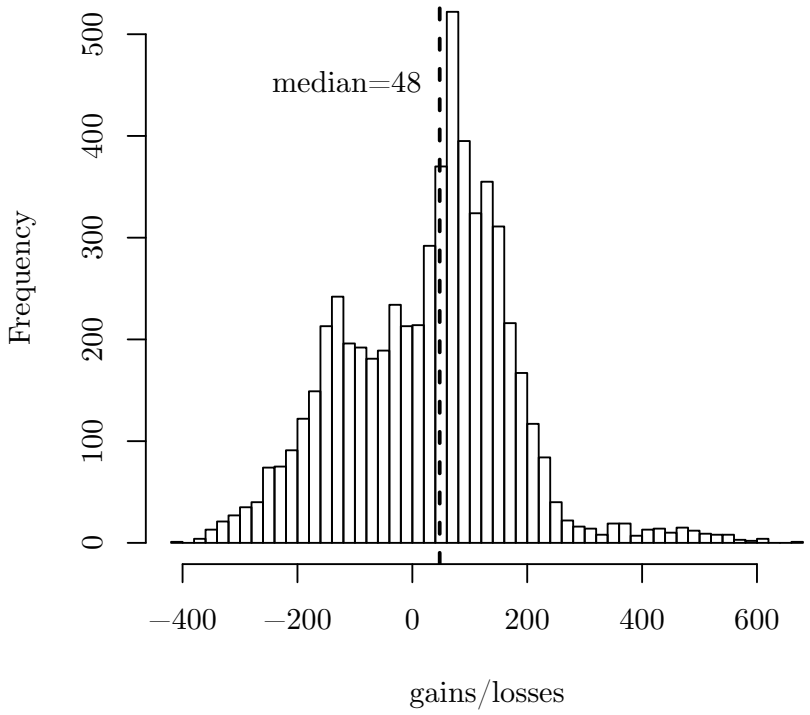


Figure 7.10: Distribution of gains and losses for the S&P500 index example. This chart shows the gain or loss for a 1 currency unit bet per point movement over a 6 month investment horizon.

that have infinitely deep pockets. Therefore, as it is, this trading strategy could not be relied upon in a real world scenario.

In this context, it is perhaps interesting to investigate how a trader's wealth would change of the course of pursuing this strategy over the 25 year period leading up to 2020. Fig. 7.11 shows the "marked-to-market" (**MTM**) value of the trader's outstanding position on each day plus accumulated realised gains or losses. The MTM value is the amount of unrealised gains or losses made each day by the trader. Clearly it is not enough for a single transaction to make money, the trader must have sufficient resources to run that position until the end of the 6 month investment horizon in order to take advantage of the accuracy of the model over that period. One can see from the plot that the trader profit/loss is moving in ranges from day to day, but these are small compared to the cumulated

Figure 7.11: Trading CFDs on the S&P500 index from 1995. In the example trading strategy, the trader buys or sells a one currency unit per point CFD on each trading day based on the prediction given by the algorithm. The trader holds this position for 130 trading days. In this way, the trader builds up net positions of up to 130 currency units per point. After the first 130 days, the trader starts closing out positions, and realised profit or losses are made. Additionally, outstanding positions are valued at the current market price. As can be seen, the performance of the strategy depends very much on when the first trade is made. Although the strategy is generally profitable, large drawdowns are certainly possible.

gains/losses over longer periods.

It is judicious that the beginning of 1995 was chosen for this example as it shows that the trader would have made very large gains for running positions of a maximum of 130 currency units at any one time, i.e., the trader is bound by the algorithm described in §7.3.1 to make one trade each day for 1 currency unit and hold this position for six months. However, if it is assumed that the trader does not have limitless wealth, and other staring periods are taken, it could easily be the situation that the trader's resources are completely depleted before he/she is able to take advantage of the predictive qualities of the algorithm.

7.3.4 Autocorrelation in time lagged data - correction of the AUC

To determine whether the algorithm described in §7.3.1, and in particular the probability measure p_i, have a predictive quality on any particular trading day, one may examine the AUC of the ROC curve for each data point. It may well be the case that on a particular day, the AUC is indicating that the classification of the predictor, p_i, is not better than a random guess, i.e. the AUC is close to 0.5. On the other hand it might be the case that the AUC is very high, and therefore, one might give more credence to a prediction on that day.

However, it is necessary to consider the effect of autocorrelation in both the predictor and response time-series and consider whether high AUC is actually a mere artefact of high autocorrelation. Clearly, in any financial market, the current price must have some correlation with the immediately preceding price. As such, one may build up a picture of the structure of how prices are correlated with prices observed over previous trading days.

In respect of the lagged percentage change in asset price, the longer the lag, the greater the autocorrelation. In the case of the probability measure, p_i, which is a function of the one data point return over the previous ω data points, a greater value of ω will increase the autocorrelation of the time series.

Since both the probability measure, p_i, and the forward price difference are auto-correlated to some degree, it would be naive to expect that even in random

simulations, the AUC for an optimally chosen value of τ will be 0.50, since there will always be window of each time-series that can be aligned to give a high level of correlation between the response and predictor which is not based on any meaningful prediction, and, as such, an elevated level of predictive quality as given by the AUC may be seen.

It is necessary, therefore, to try and quantify what value AUC can be attributed to "real" insight rather than a happy coincidence, and furthermore, what confidence limit can be ascribed to such an assertion. To illustrate the point, and to derive some comparative measures with which to benchmark results from real market data, one can examine the distribution of AUC for ROC curves resulting from randomly generated binary sequences to simulate the response variable, having been constructed to have 1-period autocorrelation similar to that of the observed data, and a sequence of randomly generated autocorrelated numbers from the interval $[0, 1]$ to simulate the predictor variable, where the random binary sequence is constructed to have a 1-period autocorrelation comparable to the that of the sequence of p_i derived from the observed data.

Firstly, a binary sequence is constructed using Bernoulli random variables. These random variables can take values of either 1 or 0. It is assumed that a value of 1 is taken if an experiment with probability ρ results in success and a 0 otherwise. Therefore to simulate the response, a binary sequence with autocorrelation is created using the following generating function

$$B_{i+1} = \begin{cases} \text{Ber}(\rho) & B_i = 1 \\ \text{Ber}(1-\rho) & B_i = 0 \end{cases} \tag{7.77}$$

The 1-period autocorrelation of this sequence is given by the covariance $\text{Cor}(B_i, B_{i+1})$. This is evaluated by firstly considering the covariance $\text{Cov}(B_i, B_{i+1}) = \text{E}[B_i B_{i+1}] - \text{E}[B_i]\,\text{E}[B_{i+1}]$. Since $P(B_i = 1) = 1/2$ implies the probability $P(B_i B_{i+1} = 1) = \rho/2$, then $\text{E}[B_i B_{i+1}] = \rho/2$. Therefore, it follows that $\text{Cov}(B_i, B_{i+1}) = (2\rho - 1)/4$. Since the standard deviation $\sigma_i = 1/2$, and $\text{Cor}(B_i, B_{i+1}) = \text{Cov}(B_i, B_{i+1})/\sigma_i \sigma_{i+1}$, the autocorrelation of the binary sequence is given by $2\rho - 1$. Therefore, if one

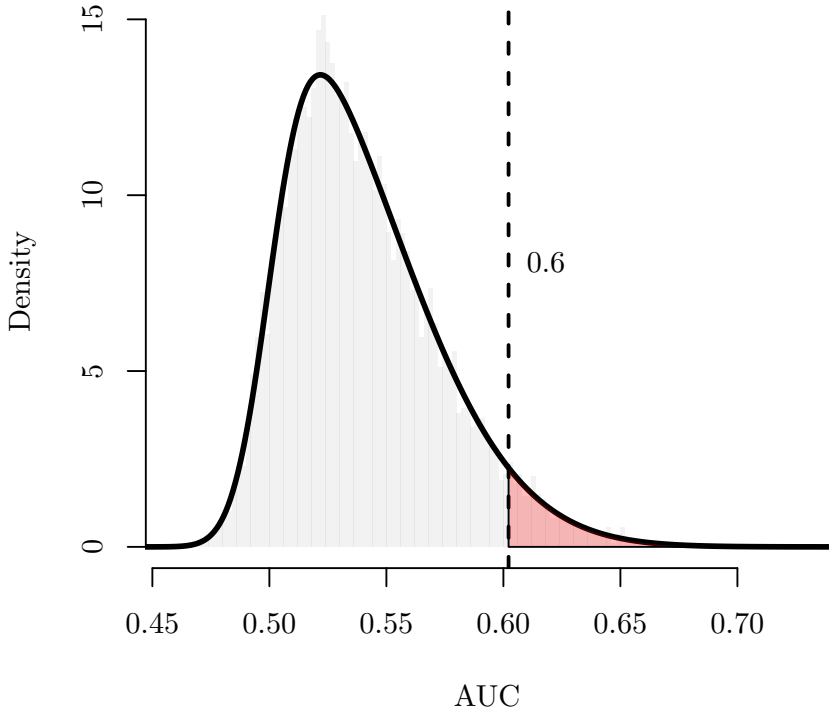


Figure 7.12: Distribution of AUC for randomly generated 5 year sequences of predictor and response which match the observed 1-period autocorrelation of the observed data of 0.88 and 0.95 respectively. As can be seen, it is certainly possible to achieve a high AUC which is purely an artefact of the autocorrelation of the sequences. The 95% quantile of this distribution is at the AUC of 0.60. Incidently, this distribution is well fitted by a skew-normal distribution

wants to generate a binary sequence with a 1-period autocorrelation of A, then one must set $\rho = (A + 1)/2$.

It is now possible to run simulations of predictor/response sequences which have similar 1-period autocorrelation to the observed data, and for each simulation, calculate the AUC of the ROC curve. This is shown in Fig. 7.12 for the 1-period autocorrelations observed in the S&P500 index example.

As can be seen in these completely random samples, there is a very large proportion of outcomes that result in a reasonable AUC. Therefore it would seem prudent to filter the trading strategy predictions by those which present and AUC in excess of the 95% quantile of the simulated distribution of AUCs. In this

		Prediction outcome		
		Higher	**Lower**	**Total**
Actual value	**Higher**	1,786 0.598	1,203 0.402	2,989
	Lower	472 0.573	352 0.427	824
	Total	2,258	1,555	

Table 7.3: Confusion matrix for predictions filtered by AUC threshold. Trading days where the AUC of the historical predictor exceeds the 95% threshold level constitute around 64% of the total number of trading days considered in this example. The table shows that the overall accuracy is reduced to 56% from 62% in the unfiltered example. Interestingly, the TPR is broadly the same as that achieved in the unfiltered example, and the reduction in accuracy is made up almost entirely in reduction of the TNR.

example the threshold AUC for filtering is 0.60.

One can see that this has quite an effect on the results of the algorithm. When looking at the resulting confusion matrix in Fig. 7.3, one can see that the accuracy has been reduced to 0.56. This would still constitute a positive result from the perspective of financial markets forecasting, but as can be seen, the true negative rate is very poor in this example.

Fig. 7.13 shows the performance of the trading strategy in the same terms as were described for the non-filtered data in Fig. 7.9. It can be clearly seen that when one considers only predictions that the AUC indicates should be unlikely due to an artefact of the inherent autocorrelation, the performance is greatly reduced. However, one may take the view that, despite this, having taken into account the autocorrelation effects, an accuracy of 0.56 would certainly indicate that the parameter p_i has an unusual predictive quality and is deserving of greater study.

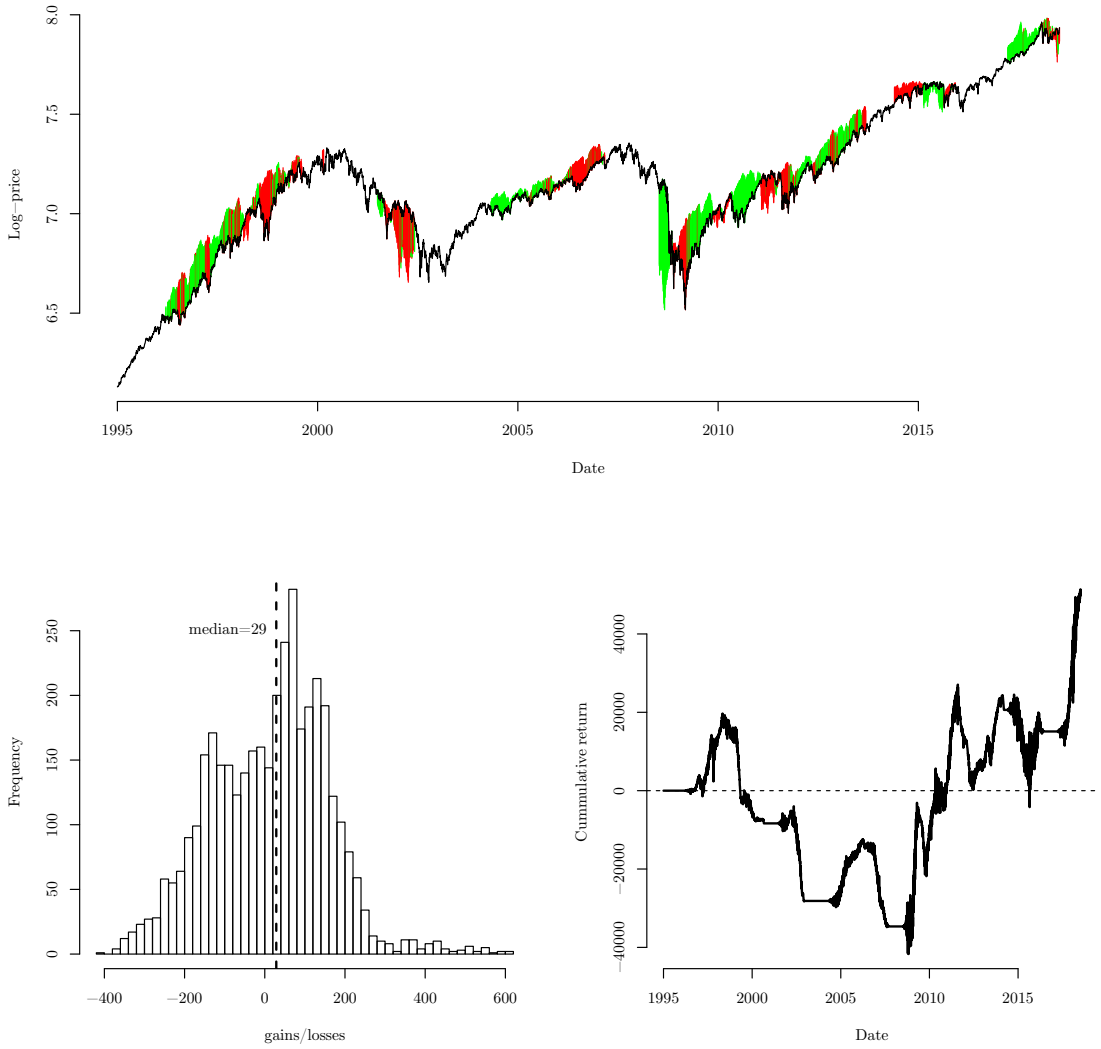


Figure 7.13: Performance of filtered predictions in respect of the S&P500 index over the 25 years preceding the end of 2019. The top graph shows the effectiveness of a 1-currency-unit bet each trading day where the AUC is in excess of the threshold AUC given by the random simulations. The bottom-left plot shows the distribution of returns, with associated median return per trade, and the bottom right plot is the daily total cash position of a trader executing the strategy from the beginning of 1995. As can be seen from the plot, although ultimately the trader can be shown to make money, he/she would need sufficient resources to weather negative returns.

Chapter 8

Longitudinal and cross-sectional study of global financial markets

8.1 Forecast results across asset classes

To test the validity or otherwise of predictions made by the algorithm described in Chapter 7, a broad study of a large number of financial time series from a variety of asset classes has been undertaken. The asset classes chosen are global stock indices, large listed equities from the UK, USA and Germany, real commodities, foreign exchange and finally the UK and US long bonds. It is important to note that in each of these studies it has been decided to use the same value for each parameter, ω, the length of historical data used at each data point, the normalisation period, ω', the minimum and maximum time period to the critical time, s_0 and s_1, and the investment horizon, τ. It would certainly be possible to solve for optimal values of these parameters at each data point in each market, but this runs the risk of over fitting a large parameter space to the data, and missing important insights that the model in a basic form may be able to give.

In this chapter, results are given for application of the algorithm over the longest data series that could be obtained easily, and freely, from sources such as Yahoo finance[1] and Quandl[2]. It should be noted that these trading strategies

[1]https://uk.finance.yahoo.com
[2]https://www.quandl.com/

have not considered trading costs or bid-offer spreads, but have taken into account the effect of daily mark-to-market of open positions.

8.1.1 Significance of prediction accuracy

As was mentioned in Chapter 7, it is important to determine whether the AUC of the ROC curve is not merely an artefact of autocorrelation in the predictor and response variables. Similarly, when one examines the accuracy of the the predictions given by the algorithm, one must devise a method to gauge the statistical significance of this prediction given the autocorrelation inherent in the predictor and response binary sequences.

For each market investigated in this chapter, the predictive algorithm produces a binary prediction of whether the market price either rises or falls over the succeeding 130 market-day period, 1 denoting a rising market, and 0 for when the market is falling. This 130-day period has been chosen arbitrarily but recognising that market prices are too volatile for short-period predictions to be accurate, and long period predictions may not be particularly suitable when using the algorithm for live trading.

When one examines these predictor binary sequences, it becomes apparent that they contain a very high degree of autocorrelation. Similarly, as discussed in the previous chapter, the response, or actual outcome of the 130 market-day period, is given by a similarly highly autocorrelated binary sequence. It would seem that before any credence is given to a prediction made in this manner, one should analyse how randomly generated predictor and response binary sequences perform in terms of predictive accuracy for a given length of sequence and 1-period autocorrelation, so that a determination can be made as to whether or not an accuracy achieved by the algorithm is a mere artefact of the inherent autocorrelation of the market price.

To make this analysis, first let the response binary sequence be denoted as A, and the predictor sequences as B. Each sequence may be given by the recurrence

relationships

$$A_{i+1} = \begin{cases} \text{Ber}(\rho_A) & A_i = 1 \\ \text{Ber}(1-\rho_A) & A_i = 0 \end{cases}, \quad B_{i+1} = \begin{cases} \text{Ber}(\rho_B) & B_i = 1 \\ \text{Ber}(1-\rho_B) & B_i = 0 \end{cases}, \tag{8.1}$$

where $i = (1, 2, \ldots, n)$ and ρ_A and ρ_B are respectively the correlation coefficients between successive entries in the binary sequences, A and B. If the success or otherwise of the prediction is given by whether or not $A_i = B_i$ then a further binary sequence, C, may be derived such that

$$C_{i+1} = \begin{cases} 1 & A_i = B_i \\ 0 & A_i \neq B_i \end{cases} \tag{8.2}$$

Since each A_{i+1} and B_{i+1} has a known correlation to A_i and B_i respectively, one can determine the resulting correlation between C_{i+1} and C_i, by considering Table. 8.1.

Predictor		**Response**		**Success**		**Probability**
A_i	A_{i+1}	B_i	B_{i+1}	C_i	C_{i+1}	$P(C_{i+1} = C_i)$
1	1	1	1	1	1	$\frac{1}{4}\rho_A\rho_B$
0	0	0	0	1	1	$\frac{1}{4}\rho_A\rho_B$
1	0	1	0	1	1	$\frac{1}{4}(1-\rho_A)(1-\rho_B)$
0	1	0	1	1	1	$\frac{1}{4}(1-\rho_A)(1-\rho_B)$
1	1	0	0	0	0	$\frac{1}{4}\rho_A\rho_B$
0	0	1	1	0	0	$\frac{1}{4}\rho_A\rho_B$
1	0	0	1	0	0	$\frac{1}{4}(1-\rho_A)(1-\rho_B)$
0	1	1	0	0	0	$\frac{1}{4}(1-\rho_A)(1-\rho_B)$

Table 8.1: 1-period lag correlation for a prediction success sequence. If binary response and predicator vectors have 1-period correlation coefficient of ρ_A and ρ_B respectively, one may find the correlation coefficient ρ_C where the i elements, C_i, of the binary vector C are given by $C_i = 1$ when $A_i = B_i$ by considering all the possible configurations of successive elements of binary vectors A and B that give rise to identical successive values of C.

By considering all possible configurations of A_i,A_{i+1},B_i and B_{i+1} that result in $C_{i+1} = C_i$, and understanding that in isolation the probability $P(A_i = 1) =$

$P(B_i = 1) = 1/2$, it is easily demonstrated that

$$P(C_{i+1} = C_i) = 2\rho_A\rho_B - \rho_A - \rho_B + 1 = \rho_C \,. \tag{8.3}$$

Therefore, given ρ_A and ρ_B, one may model the prediction success of the binary sequence by the recursive relationship

$$C_{i+1} = \begin{cases} \text{Ber}(\rho_C) & C_i = 1 \\ \text{Ber}(1-\rho_C) & C_i = 0 \end{cases} \tag{8.4}$$

From this relationship, it is interesting to investigate the probability distribution of successes, or more precisely the probability distribution of the implied accuracy of the success sequence. For the purposes of this chapter, the set of accuracy measures for N sequences is denoted as Θ.

The vector C can be considered as a sequence of Bernoulli trials, i.e., trials where there is a binary outcome. It is generally known that the probability distribution of the sum of successes in a sequence of independent Bernoulli trials is described by a binomial distribution. However, as shown in Gonzalez-Barrios (1998), the sum of non-independent Bernoulli trials can be approximated by a binomial distribution. Therefore, in this thesis, an assumption is made that $\sum_i^n C_i$ is drawn from a binomial distribution, and as is usual, the beta distribution is used to describe the distribution of the probability of success. As such, at least for large n, the probability distribution of Θ should be described well by the beta distribution whereby, given the shape parameters α and β, the probability density function is written as

$$f(x;\alpha,\beta) = \frac{1}{\text{B}(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} \tag{8.5}$$

where B is the beta function, defined as

$$\text{B}(\alpha,\beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt \,. \tag{8.6}$$

If it is reasonable to assume a beta distribution, the expectation $\mathrm{E}[\Theta]$ is given by

$$\begin{aligned}\mathrm{E}[\Theta] &= \int_0^1 x \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha,\beta)} dx \\ &= \frac{\mathrm{B}(\alpha+1,\beta)}{\mathrm{B}(\alpha,\beta)}\,. \end{aligned} \tag{8.7}$$

However, since the beta function may also be defined in terms of the gamma function

$$\mathrm{B}(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\,, \tag{8.8}$$

one may write

$$\begin{aligned}\mathrm{E}[\Theta] &= \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \\ &= \frac{\Gamma(\alpha+1)\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+1)\Gamma(\alpha)}\,. \end{aligned} \tag{8.9}$$

Since $\Gamma(z+1) = z\Gamma(z)$, one can readily see that

$$\mathrm{E}[\Theta] = \frac{\alpha}{\alpha+\beta}\,. \tag{8.10}$$

However, in the case at hand, it is clear that $\mathrm{E}[C] = 1/2$ giving the necessary relationship $\alpha = \beta$. Similarly, the variance $\mathrm{Var}(\Theta)$ is given by

$$\begin{aligned}\mathrm{Var}(\Theta) &= \mathrm{E}\big[\Theta^2\big] - \mathrm{E}[\Theta]^2 \\ &= \int_0^1 x^2 \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha,\beta)} dx - \left(\frac{\alpha}{\alpha+\beta}\right)^2 \\ &= \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} - \left(\frac{\alpha}{\alpha+\beta}\right)^2 \\ &= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \end{aligned} \tag{8.11}$$

but since in this case $\alpha = \beta$, the variance is given by

$$\mathrm{Var}(\Theta) = \frac{1}{4(2\beta+1)} \tag{8.12}$$

All that remains to describe the probability density function of the accuracy

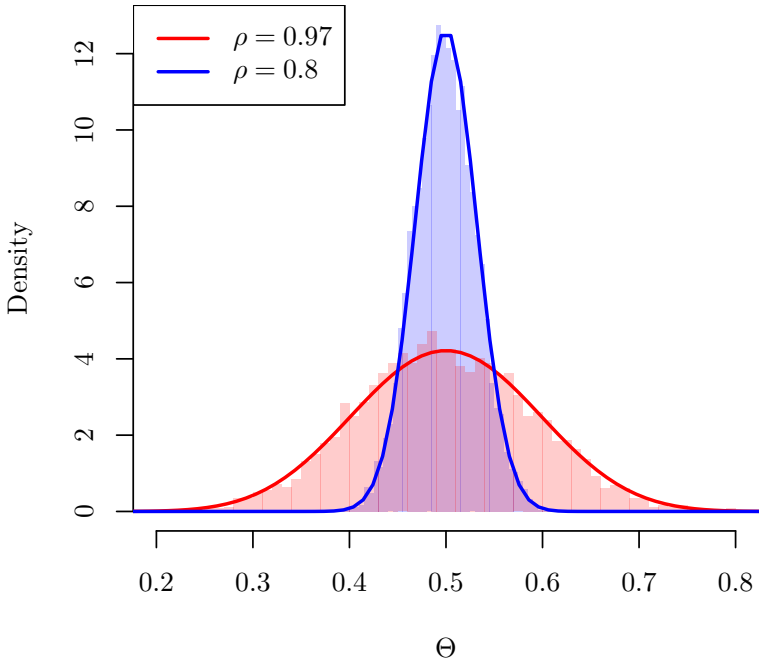


Figure 8.1: For binary sequences of 1000 data points, the distribution of the accuracy is shown for two values of ρ_C, and the associated beta distribution derived from the expression for the shape parameter. From a visual inspection, it would appear that the fitted distribution correlates very well to the experimental data. This demonstrates quite clearly the fact that as the autocorrelation increases, the greater the variance, and the more likely higher accuracy measures are merely an artefact of the autocorrelation.

metric, is to derive an expression for the variance of this measure, $\mathrm{Var}(\Theta)$ in terms of the correlation coefficient, ρ_C and the length of the sequence, n. The variance of Θ is given by

$$\mathrm{Var}(\Theta) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{Cov}(C_i, C_j)\,. \tag{8.13}$$

Since the covariance $\mathrm{Cov}(C_i, C_j)$ is given by

$$\begin{aligned} \mathrm{Cov}(C_i, C_j) &= \mathrm{E}[C_i C_j] - \mathrm{E}[C_i]\mathrm{E}[C_j] \\ &= \mathrm{E}[C_i C_j] - \frac{1}{4}\,, \end{aligned} \tag{8.14}$$

one must find an expression for $\mathrm{E}[C_i C_j]$. The product $C_i C_j$ is 1 when $C_i = C_j = 1$

and zero otherwise, so $C_i = C_j = 1$ is the only case which needs to be considered. Therefore, in this case, the sequence $(C_i, C_{i+1}, \ldots, C_{j-1}, C_j)$ where $i < j$ and $C_i = C_j = 1$, becomes $(1, C_{i+1}, \ldots, C_{j-1}, 1)$. In this sequence there can be a maximum of $j - i$ changes of state (i.e., from 1 to 0), and, since the states of C_i and C_j are the same, the number of changes of states must be even. Therefore, letting $m = |j - i|$,

$$\begin{aligned} \mathrm{E}[C_i C_j] &= \frac{1}{2}\left(\frac{1}{2}\sum_{k=0}^{m}\binom{m}{k}(1-\rho_C)^k\rho_C^{m-k} + \frac{1}{2}\sum_{k=0}^{m}\binom{m}{k}(1-\rho_C^k\rho_C^{m-k}(-1)^k\right) \\ &= \frac{1}{4}\left(1 + (2\rho_C - 1)^m\right) . \end{aligned} \tag{8.15}$$

It follows that

$$\mathrm{Cov}[C_i C_j] = \frac{1}{4}(2\rho_C - 1)^m . \tag{8.16}$$

Therefore, the variance of Θ is given by

$$\mathrm{Var}(\Theta) = \frac{1}{4n^2}\sum_{m}(2\rho_C - 1)^m \tag{8.17}$$

where the sum is taken over all possible values of $m = |j - i|$ where $j = 1, 2 \ldots, n$ and $i = 1, 2, \ldots, n$. The possible values of m are can be shown in the following matrix

$$m_{i,j} = \begin{bmatrix} 0 & 1 & 2 & \ldots & n-1 \\ 1 & 0 & 1 & \ldots & n-2 \\ 2 & 1 & 0 & \ldots & n-2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n-1 & n-2 & n-3 & \ldots & 0 \end{bmatrix} \tag{8.18}$$

The values contained in this matrix are comprised of the sequence $k = (0 \ldots n-1)$ with each element of k occurring $2(n - k)$ times except the first element which

occurs only n times. As such, one may express the variance as

$$\begin{aligned}\mathrm{Var}(\Theta) &= \frac{1}{4n} + \frac{1}{2n^2}\sum_{k=1}^{n-1}(n-k)\,(2\rho_C - 1)^k \\ &= \frac{1}{4n} + \frac{1}{2n}\sum_{k=1}^{n-1}(2\rho_C - 1)^k - \frac{1}{2n^2}\sum_{k=1}^{n-1}k\,(2\rho_C - 1)^k \end{aligned} \tag{8.19}$$

Letting $\hat{\rho} = (2\rho_C - 1)$, from known finite sums, one may write

$$\mathrm{Var}(\Theta) = \frac{1}{4n} + \frac{\hat{\rho}^n - \hat{\rho}}{2n(\hat{\rho} - 1)} + \frac{(n-1)\hat{\rho}^{n+1} - n\hat{\rho}^n + \hat{\rho}}{2n^2(\hat{\rho} - 1)^2}\,. \tag{8.20}$$

Therefore, for n Bernoulli trials where successive trials are correlated with the previous trial with correlation coefficient ρ_C, the percentage success rate, or accuracy, can be described by a symmetrical beta distribution with shape parameter

$$\beta = \frac{1}{8\mathrm{Var}(\Theta)} - \frac{1}{2}\,. \tag{8.21}$$

Visual inspection of Fig. 8.1 shows the high quality of the fit of the beta distribution to the empirical data given two values of ρ_C. The plot highlights the concern that as autocorrelation in response and predictor binary sequences increases, the greater the variance in the accuracy of the prediction, and the more likely it is that accuracies that would on first inspection seem impressive, may be due solely to the high level of autocorrelation.

This expression was tested against a data set of $N = 200$ correlated Bernoulli trial sequences, where n was chosen from the interval $(500, 2000)$ data points, and the value of ρ_C was sampled from the uniform distribution $U(0, 1)$. Having simulated each trial 500 times, to test the veracity of the model, randomly chosen empirical quantiles, sampled from $U(0, 1)$, were compared to that same quantile derived from the beta distribution with the calculated shape parameters. The r.m.s. error for this test was shown to be 0.0042. It is clear that the approximation given in this section is sufficient to gauge a measure of the confidence limit for the accuracy apparently achieved by the predictive algorithm, which may indicate how much credence should be given to any such accuracy.

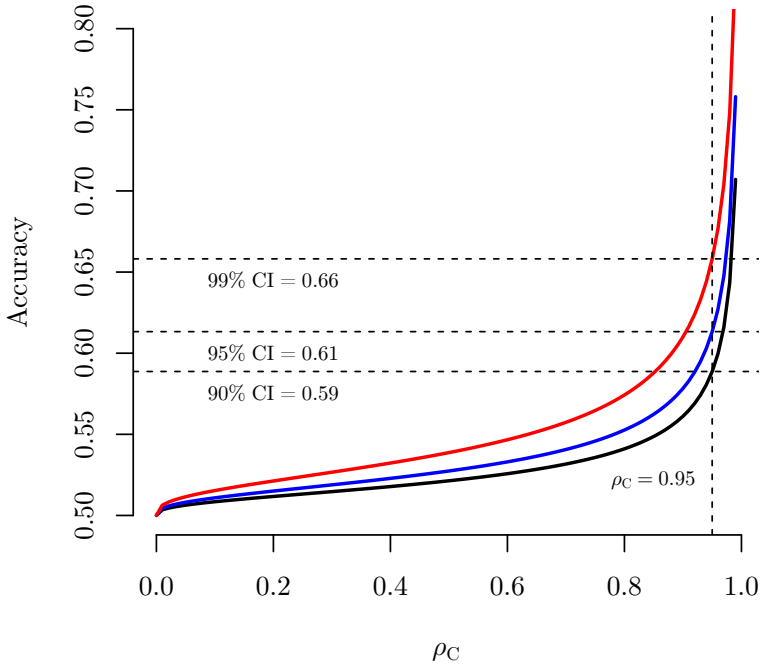


Figure 8.2: Accuracy confidence limits for a 1000 data point binary sequence. In analysis of financial time series, the value of ρ_C is very often in excess of 0.95. At this level, once can see that to be confident at the 99% level that the prediction has not merely arise by chance, the accuracy or the predictions would have to be in excess of 0.66. By any measure this is difficult to achieve in financial time series forecasting.

Finally in this section, it is worth quantifying those accuracies at which one may feel are unlikely to be due to the autocorrelation of the response and predictor. Fig.8.2 shows those accuracies that correspond to confidence limits of 90%, 95%, and 99% for a given ρ_C and $n = 1000$. As can be seen, if an investigation of a particular financial time series yields a value of $\rho_C = 0.95$ derived from the 1-period autocorrelation of the predictor and response, then once may be reasonably confident that an accuracy in excess would be unlikely an artefact of autocorrelation. On the other hand, if the calculated accuracy was less than 0.59, in of itself a very high accuracy in respect of prediction financial markets, one may feel that there is too much of a risk that there is no useful information contained in these predictions.

8.2 Trading strategies

For the purposes of analysing a broad cross section of financial markets, and to determine whether or not the algorithm described in §7.3.1 provides a workable trading methodology, one must decide how to execute transactions given the calculated prediction. In the course of this research, various strategies were considered based on the absolute value of $\bar{\nu}'$. Initially, it was considered that, when a buy signal was given by the parameter breaching a threshold value, the fictitious trader would enter into a long position and only reverse this position once a sell signal had been given by the parameter breaching another threshold value. This applies in the context of buying a physical equity, i.e. purchasing actual shares in a company. If this is the case one must also consider the cost of entering into such a transaction, including funding the capital required for the initial purchase and the bid-offer spread (the difference between where one may purchase and sell an equity). Furthermore, it is unrealistic, in the context of this strategy, to consider the trader being in a position to short equities he/she does not already own. In financial markets this is clearly possible, but one must consider the cost of borrowing the stock to cover the short position. This cost is difficult to accurately estimate, can move quite considerably, particularly over dividend dates, and there is no historical data on this matter. Additionally, by using this trading strategy, it is not straightforward to gauge the accuracy, and practical usefulness, of a particular prediction as outcomes are very sensitive to when the fictitious trader decides to start employing the strategy.

It is for this reason that a strategy has been chosen that enables one to analyse each prediction in isolation, and which can be used by traders in real-world trading environments. This is the method that has been described briefly in Chapter 7. In this strategy, the current value of $\bar{\nu}'$ and its historical values are used to give binary predictions of whether the market will be higher or lower after a particular time has elapsed. As has been mentioned before, there may be ways to continually update the amount of historical data used, the length or the normalisation period and the investment horizon to give optimal historical accuracy, and then apply these updated metrics to live predictions. However, as a first step, it seems

sensible to fix these parameters at arbitrary (although justifiable) levels such that a baseline of effectiveness can be gauged that is agnostic to the peculiarities of any particular market. Therefore, for this chapter, the research is limited to four simple strategies, which go some way to providing a framework to judge the utility of detecting discrete scale invariance in this manner.

Contract for differences As described in Chapter 7, it is proposed that a fictitious trader enters into long or short contract for differences (**CFD**) over the target asset. By using CFDs, the trader is taking a "leveraged" position in that he/she has not been required to fund a purchase of the underlying asset, but has essentially been extended finance by the exchange to be granted exposure to the price movement of the underlying asset in a amount equivalent to holding a particular asset. For example, consider a trader who decides to go long (i.e. buy) Apple Inc. CFDs for US$10 per point at the current price of $280 per share. This means for each price point the stock rises the trader is paid $10 by the exchange. Conversely, if the price falls, the trader loses $10. To have the same upside exposure in the physical market by buying the underlying stocks outright, the trader would need to spend $2800 of his/her cash, and have the opportunity cost of earning risk-free interest on this amount. However, in the CFD market, the trader (if a professional) will only need to deposit around 0.45% of this amount to the exchange to cover the risk of loss. This is called the initial margin and must be kept in the trader's account with the exchange while the position is open. As the market moves the trader may be subject to a margin call by the exchange. This happens when the market movement results in unrealised losses on the trading position, and the value of the deposit falling. The amount of margin call is referred to as the maintenance, or variation margin. As can be seen, although it may be cheap for the trader to open a position, as the market moves against the trade, the trader must be in a position to fund any margin call, otherwise the exchange will "close out" the position automatically. Clearly if this happens in the context of the proposed algorithm, market movements during the investment horizon may give rise to margin calls which are unable to be funded.

Therefore it is important to monitor the size of drawdowns suffered while the strategy progresses.

This strategy is the basic strategy that has been used for the purposes of this analysis, and it has been assumed that there are no bid/offer spreads and that the trader has enough resources to cover any margin call. Additionally, the funding of these margin calls have not been taken into account. These assumptions are not realistic, but the overall construct serves as a fair approximation of what may be possible, and enables one to get a good first view of the performance of the algorithm.

Filtered CFD The CFD strategy assumes that the trader enters into a position every day based on the algorithm's (§7.3.1) prediction. However, it has been shown in Chapter 7 that the AUC of the ROC curve for the predictor and response can certainly be merely an artefact of the autocorrelation of these time series. Therefore, one could imagine, that where the historical AUD of the prediction falls below a certain value, the trader decides not to enter the market based on this predication. In the following analysis, the AUC threshold value has been taken to be the 95% quantile of the AUC distribution from a simulation of randomly generated response and predictor time-series. Therefore, if the AUC threshold is not breached on any trading day, the trader will not enter into a CFD transaction. It is hoped that, in this way, only those transaction likely to be predicted from real insights are acted upon.

Predicted long The construction of what constitutes an anti-bubble in the preceding analysis is open to challenge. The interpretation of this thesis diverges from the position taken in the literature, in that an assumption has been made in Chapters 7 and 8 that an anti-bubble is essentially the opposite of a bubble, with superexponential drawdowns decorated with log periodic oscillations with increasing frequency approaching a critical time, t_c. As has been described earlier in this thesis, and in particular in Chapter 5, the traditional model of an anti-bubble is one where the critical time, t_c, occurs prior to the the start of the anti-bubble regime, and, at least in the log-periodic models of Sornette et. al.,

the amplitude of the oscillations saturate as the time from t_c increases.

The idea behind the trading strategies above is that when there are signals that a bubble is forming the traders should buy into that bubble, and conversely should keep selling into an anti-bubble. In order to see whether the prediction results are equally effective in both bubble and anti-bubble regimes, a strategy is considered whereby the trader only enters into transactions based on predictions that forecast a rise in the market prices, effectively filtering out any transaction based on the detection of an anti-bubble. It will be seen later that this provides a very interesting comparison to the other strategies.

Long only In order to have a baseline to compare to other strategies that use the $\bar{\nu}'$ measure, a "control" strategy has been incorporated into the research which considers a naive trader who, each trading day, simply enters into a buy transaction which is held for six months regardless of how the market develops over time. Over the long term this would have been a very judicious strategy since stock and commodity markets have generally seen steady price inflation, and, as such, just buying has been a profitable position to have taken. Therefore, it is considered that this strategy is a reasonable target over which any other trading strategy should aim to outperform, particularly since transactions costs have not been taken into account.

8.3 Results

The above strategies have been applied to four asset classes, namely stock indices, common stocks, commodities, and foreign exchange, and a total of 111 underlying financial instruments. The results have been separated by asset class and the confusion matrix for both the *CFD* strategy and the *filtered CFD* strategy have been determined. Additionally, the probability is given of achieving the observed accuracy of the *CFD* strategy if the response and predictor binary series had been generated randomly as a correlated sequence of Bernoulli trails based the observed 1-period autocorrelation of the actual predictor and response, and the 95% confidence interval of the accuracy for such probability is given for that

particular market. Similarly, the 95% confidence interval is given for the AUC of the ROC curve such that one may determine which transactions should be included in the filtered *CFD* strategy.

In analysing these results, it is important to report which underlying assets are more accurately predicted. However, accuracy does not tell the whole story, since its value does not distinguish between the TPR and TNR, which as will be seen, can be significantly different. Therefore, although the results are ordered by accuracy of the *CFD* strategy, those markets are highlighted where both the true positive rate and the true negative rate of this strategy are both either 50% or more. In these case, there would seem there is the possibility that the strategy can be universal applied. Where it is the case that only the true positive rate is in excess on 50% it may be that the *predicted long* strategy is more suitable. It is also important to note in which markets does the act of using the *filtered CFD* strategy enhance the performance of the prediction.

Clearly, the accuracy of a binary prediction must be examined in the context of the absolute amount of gains and losses achieved, i.e. a trading strategy that gives accurate binary higher or lower predictions is of no use if those losses made by inaccurate prediction are consistently larger than any gains made by accurate predictions. It is convenient that the *CFD* trading strategy that has been chosen allows for each transaction to be examined in isolation, and aggregated with all other transactions to derive a distribution of returns. The results report the mean return, expressed as a percentage of the market price when the position was opened, over the whole time series for each of the trading strategies mentioned above, and the best performing strategy is highlighted.

Finally, the *filtered CFD*, *predicted long* and *long only* strategies are compared with the *CFD* strategy in terms of the maximum gains and losses. Since CFD trading is very risky due to the consequences of potential being closed out, it is certainly desirable that any trading strategy mitigates this risk as far as possible. Therefore, the results report the maximum gain and losses of each strategy compared to that of the CFD strategy. In this way one can determine whether there is a benefit to be gained by switching to one of the alternatives.

		CFD			Filtered CFD				95% CI	
Asset	**Country**	**ACCR**	**TPR**	**TNR**	**ACCR.F**	**TPR.F**	**TNR.F**	**P(>ACCR)**	**ACCR**	**AUC**
DAX	Germany	0.65	0.67	0.62	0.63	0.68	0.48	0.00	0.54	0.61
CAC 40	France	0.64	0.66	0.61	0.67	0.71	0.61	0.00	0.54	0.60
S&P 500	USA	0.64	0.66	0.59	0.58	0.63	0.46	0.00	0.53	0.60
Jakarta Composite	Indonesia	0.64	0.70	0.43	0.65	0.73	0.33	0.00	0.55	0.61
ESTX 50	Europe	0.63	0.67	0.58	0.61	0.66	0.55	0.00	0.54	0.61
FTSE 500	UK	0.63	0.67	0.56	0.65	0.70	0.53	0.00	0.53	0.59
Nasdaq Composite	USA	0.63	0.66	0.57	0.55	0.60	0.46	0.00	0.53	0.60
NYSE Composite	USA	0.63	0.66	0.55	0.60	0.66	0.44	0.00	0.53	0.60
ASX 200	USA	0.62	0.68	0.48	0.59	0.64	0.50	0.00	0.54	0.60
NYSE AMEX Composite	USA	0.61	0.66	0.52	0.61	0.72	0.29	0.00	0.54	0.60
TSX Composite	Canada	0.61	0.66	0.50	0.60	0.73	0.13	0.00	0.54	0.60
TSEC Weighted	Taiwan	0.60	0.65	0.50	0.75	0.81	0.66	0.00	0.54	0.60
Merval	Argentina	0.59	0.61	0.54	0.56	0.56	0.57	0.00	0.54	0.61
Tel Aviv	Israel	0.59	0.61	0.54	0.56	0.56	0.57	0.00	0.54	0.61
Dow Jones	USA	0.59	0.62	0.52	0.56	0.60	0.49	0.00	0.53	0.59
All Ordinaries	Australia	0.59	0.64	0.49	0.57	0.59	0.52	0.00	0.53	0.60
KOSPI Composite	South Korea	0.59	0.66	0.46	0.62	0.72	0.38	0.00	0.54	0.60
BSE SNESEX	India	0.59	0.62	0.49	0.72	0.76	0.60	0.00	0.55	0.61
SSE Composite	China	0.56	0.64	0.44	0.57	0.64	0.48	0.00	0.54	0.61
Hang Seng	Hong Kong	0.56	0.64	0.44	0.57	0.64	0.48	0.00	0.54	0.61
Russell 2000	USA	0.56	0.60	0.48	0.52	0.51	0.53	0.00	0.53	0.59
IPSA	Chile	0.56	0.74	0.33	0.68	0.90	0.27	0.05	0.56	0.62
Bursa Malaysia KLCI	Malaysia	0.55	0.65	0.40	0.59	0.62	0.55	0.04	0.55	0.62
BEL 20	Belgium	0.54	0.57	0.50	0.59	0.59	0.59	0.04	0.54	0.61
MOEX Russia	Russia	0.54	0.57	0.50	0.59	0.60	0.59	0.04	0.54	0.61
STI	Singapore	0.54	0.65	0.39	0.54	0.61	0.43	0.06	0.54	0.63
IPC Mexico	Mexico	0.52	0.60	0.39	0.58	0.65	0.48	0.18	0.54	0.60
Euronext 100	Europe	0.52	0.64	0.30	0.51	0.75	0.08	0.24	0.55	0.60
Nikkei 225	Japan	0.51	0.59	0.43	0.55	0.52	0.58	0.33	0.54	0.61
NZ50	New Zealand	0.46	0.53	0.15	0.44	0.44	0.41	0.84	0.57	0.63
Ibovespa	Brazil	0.41	0.52	0.24	0.40	0.36	0.51	1.00	0.54	0.60

Table 8.2: Results of a broad selection of global stock market indices. The algorithm described in §7.3.1 was applied to time-series of around 20-25 years in length. The blue shaded areas are indicative of when both the TPR and TNR are in excess of 50% for the *CFD* strategy. Those areas highlighted in green, indicate where the same is true for the *filtered CFD* strategy.

8.3.1 Global stock indices

The first asset class to be examined is global stock indices. Table 8.2 shows the confusion matrix for the *CFD* and *filtered CFD* strategies, along with the probability that the calculated prediction accuracy would have been achieved if the response and predictor had been generated randomly with the observed 1-period autocorrelation. Additionally are shown the 95% confidence intervals of this probability and the AUC of the ROC curve of the response and predictor, if they had been generated similarly randomly.

It can be seen from the above table that the majority of these indices yield accuracy figures for the *CFD* strategy in excess of 50%. Encouragingly the accuracy figures obtained are, save for the six markets which have performed the worst in terms of the accuracy for the *CFD* strategy, sufficiently unlikely to have

Asset	Country	Mean return CFD	Filtered CFD	Pred. long	Long only	Maximum Loss (as % of CFD max loss) Filtered	Pred. long	Long only	Maximum Gain (as % of CFD max gain) Filtered	Pred. long	Long only
Merval	Argentina	7.58	4.55	22.62	18.37	106	18	21	82	217	416
Tel Aviv	Israel	7.58	4.57	22.62	18.37	105	18	21	82	217	416
Jakarta Composite	Indonesia	6.89	5.73	13.19	10.80	74	57	67	71	119	149
Nasdaq Composite	USA	5.37	4.98	9.64	6.05	76	82	269	49	103	104
BSE SNESEX	India	3.97	8.06	11.88	10.15	84	41	137	78	162	255
NYSE AMEX Composite	USA	3.66	3.66	6.49	4.13	63	51	162	54	91	87
KOSPI Composite	South Korea	3.64	4.20	7.10	5.06	70	69	113	63	106	127
S&P 500	USA	3.53	1.89	6.56	4.18	70	64	162	46	109	115
CAC 40	France	3.51	4.28	6.06	3.38	66	62	212	78	65	58
ESTX 50	Europe	3.46	2.56	5.66	2.94	91	82	255	51	65	67
DAX	Germany	3.29	1.96	7.21	5.02	86	49	137	58	100	126
NYSE Composite	USA	3.17	2.73	5.73	3.65	98	92	294	57	97	91
FTSE 500	UK	2.85	3.79	4.24	2.12	99	155	318	74	79	56
IPSA	Chile	2.84	6.74	4.04	2.91	61	102	90	65	102	106
TSX Composite	Canada	2.55	1.46	5.01	3.55	117	89	167	36	99	101
ASX 200	USA	2.53	2.18	4.33	2.96	100	71	310	62	84	80
Dow Jones	USA	2.38	0.79	5.75	4.26	81	64	151	32	136	167
TSEC Weighted	Taiwan	2.36	8.21	5.21	3.90	39	56	228	58	96	97
BEL 20	Belgium	2.04	2.70	4.13	2.48	93	101	197	91	65	82
MOEX Russia	Russia	2.04	2.75	4.13	2.48	96	101	197	91	65	82
All Ordinaries	Australia	1.98	1.90	4.13	2.95	100	60	293	69	91	94
SSE Composite	China	1.82	2.34	4.75	3.92	50	82	128	90	111	131
Hang Seng	Hong Kong	1.82	2.33	4.75	3.92	50	82	128	89	111	131
IPC Mexico	Mexico	1.68	3.44	6.89	6.59	55	46	105	94	133	217
Russell 2000	USA	1.57	-0.01	5.35	4.57	87	60	157	60	143	189
Nikkei 225	Japan	1.53	0.57	2.74	1.67	92	100	161	43	65	81
STI	Singapore	1.20	1.51	2.88	2.46	65	101	135	62	92	90
Bursa Malaysia KLCI	Malaysia	0.30	0.63	2.45	2.80	44	57	77	67	108	146
NZ50	New Zealand	-0.27	-1.04	5.40	6.65	87	11	7	61	135	315
Ibovespa	Brazil	-0.56	-3.89	6.57	8.53	59	65	54	92	118	279
Euronext 100	Europe	-0.81	-1.21	0.97	2.08	58	89	98	60	101	184

Table 8.3: Trading results for global stock indices. This table shows the mean return for each strategy in each market expressed as a percentage of the asset price when a trading position was opened. The table also shows a comparison, in percentage terms, of the maximum gain and loss between each strategy and the *CFD* strategy.

been achieved by virtue of the autocorrelation alone. However, less than half of these experiments have resulted in TPR and TNR both being over 50%. The cases where both TPR and TNR are in excess of 50% are highlighted in blue.

Turning now to the *filtered CFD* transactions, whereby the fictitious trader only executes transactions when the historical AUC of the predictions is in excess of the 95% confidence interval described in Chapter 7. Those markets where both the TNR and TPR are in excess of 50% are highlighted in green. Table 8.2 shows a significant improvement in some of the worst performing markets, and several of the better performing markets. Overall, it would seem that switching to the *filtered CFD* strategy would considerably enhance the confusion matrix of the resulting predictions.

The trading results for stock indices are shown in the Table 8.3. One can see that the rank of each market in terms of accuracy of the *CFD* strategy does not equal the rank of mean percentage returns. The general theme is that the *CFD*

strategy on average provides strongly positive returns. However, disappointingly, in no market is this mean return better than if the trader would have blindly gone long on every trading day. Towards the lower end of the table, mean percentage returns are generally enhanced by the *filtered CFD* strategy, but the effect is sporadic. However, there is a very clear enhancement given by using the *predicted long* strategy. In nearly all cases this strategy outperforms all other strategies by some margin. Furthermore, this strategy reduces the maximum loss significantly when compared to the *long only* strategy. In general the *long only* strategy would have very large swings in profitability, increasing the risk that a trader in the real world would be stopped out. This issue seems quite well mitigated by both the *predicted long* and *filtered CFD* strategies. However, the *filtered CFD* strategy mitigates this risk by giving up on potential gains. The *predicted long* strategy does not suffer so much from this trade off.

This result may be expected in that stock indices have been on an upward trajectory for many years, but it is very encouraging that the *predicted long* strategy has such strong performance in respect of returns and mitigation of large drawdown risk. It also may indicate that the algorithm is really only successful in predicting discrete scale invariance in respect of asset price bubbles, and that the downside predictions are based on flawed assumptions in respect of the mechanics of anti-bubbles.

8.3.2 UK, US and German equities

Analysis of individual stocks from the UK, US and German markets tells a very different story in terms of the performance of the strategies. In all jurisdictions it is apparent that those stocks that have resulted in strong TPR and TNR are much rarer than has been witnessed in stock index markets. Similarly, enhancement on TPR and TNR performance by switching to the *filtered CFD* strategy does not seem to be a generally advantageous route to take. In particular, enhancements are totally absent in the US market, and fairly sparse in the UK and German markets. This is not to say the accuracy figures are not more likely to be due to the autocorrelations, but that the TNR seems consistently poor.

This poor performance is reflected in the table of returns across all three jurisdictions with the *CFD* strategy performing much worse than *long only*. Interestingly, in these markets the *filtered CFD* strategy very much underperforms even the *CFD* strategy, but again the winning strategy seems to be *predicted long*, particularly in Germany and the UK.

In terms of risk mitigation, it would appear that in general both the *filtered CFD* and *predicted long* strategies provide a good level of downside protection, but again the *predicted long* strategy gives up much less in terms of upside.

Table of results – UK equities

		CFD			Filtered CFD				95% CI	
Asset	**Country**	**ACCR**	**TPR**	**TNR**	**ACCR.F**	**TPR.F**	**TNR.F**	**P(>ACCR)**	**ACCR**	**AUC**
St. James's Place plc	UK	0.62	0.65	0.57	0.59	0.62	0.53	0.00	0.53	0.60
Centrica plc	UK	0.62	0.73	0.49	0.70	0.73	0.65	0.00	0.54	0.60
Ashtead Group plc	UK	0.62	0.65	0.56	0.59	0.60	0.58	0.00	0.54	0.62
Rentokil Initial plc	UK	0.60	0.62	0.54	0.53	0.52	0.55	0.00	0.54	0.60
British American Tobacco plc	UK	0.60	0.62	0.51	0.65	0.71	0.51	0.00	0.53	0.59
Prudential plc	UK	0.59	0.63	0.53	0.54	0.60	0.45	0.00	0.53	0.58
Vodafone Group plc	UK	0.58	0.59	0.56	0.53	0.59	0.45	0.00	0.53	0.60
Tesco plc	UK	0.57	0.64	0.47	0.53	0.50	0.57	0.00	0.53	0.58
Rolls-Royce Holdings plc	UK	0.56	0.64	0.42	0.61	0.71	0.43	0.00	0.53	0.60
Meggitt plc	UK	0.55	0.60	0.44	0.55	0.63	0.36	0.01	0.53	0.60
Compass Group plc	UK	0.55	0.60	0.33	0.50	0.58	0.15	0.02	0.54	0.59
Associated British Foods plc	UK	0.54	0.56	0.48	0.49	0.45	0.57	0.01	0.53	0.59
Smiths Group plc	UK	0.54	0.61	0.43	0.45	0.56	0.29	0.02	0.53	0.60
Schroders plc	UK	0.53	0.59	0.45	0.56	0.75	0.27	0.09	0.54	0.61
Royal Dutch Shell plc	UK	0.52	0.62	0.36	0.52	0.54	0.49	0.20	0.54	0.59
Antofagasta plc	UK	0.52	0.58	0.40	0.48	0.50	0.45	0.16	0.53	0.60
Cello Health plc	UK	0.51	0.60	0.40	0.51	0.88	0.21	0.35	0.54	0.59
BAE Systems plc	UK	0.51	0.57	0.41	0.49	0.51	0.47	0.31	0.53	0.60
Reckitt Benckiser Group plc	UK	0.49	0.52	0.43	0.47	0.47	0.49	0.73	0.53	0.58
Persimmon plc	UK	0.47	0.52	0.37	0.47	0.50	0.38	0.92	0.53	0.60
Spirax-Sarco Engineering plc	UK	0.43	0.50	0.28	0.46	0.47	0.37	1.00	0.53	0.60

Table 8.4: UK equity markets prediction performance. These are the results of a selection of individual stocks from the FTSE100 index. As before, the algorithm described in §7.3.1 was applied to time-series of around 20-25 years in length. One can see that there is a portion of the selected stocks where both the TPR and TNR are in excess of 50% for the *CFD* strategy, but the majority do not fulfil this criteria. The prediction success is not enhanced by using the *filtered CFD* strategy.

		Mean return				Maximum Loss (as % of CFD max loss)			Maximum Gain (as % of CFD max gain)		
Asset	**Country**	**CFD**	**Filtered CFD**	**Pred. long**	**Long only**	**Filtered**	**Pred. long**	**Long only**	**Filtered**	**Pred. long**	**Long only**
Persimmon plc	UK	12.25	27.40	25.51	16.60	24	105	140	56	65	58
British American Tobacco plc	UK	9.25	5.47	22.31	18.33	86	65	77	24	166	239
Prudential plc	UK	7.04	11.61	14.55	11.73	32	63	118	63	103	140
Rentokil Initial plc	UK	4.76	-3.97	13.25	12.25	67	76	78	28	113	165
St. James's Place plc	UK	4.60	5.13	9.22	7.89	66	52	66	75	122	163
Tesco plc	UK	3.83	-2.09	9.87	7.41	68	136	268	29	71	61
Royal Dutch Shell plc	UK	3.04	3.18	7.48	5.99	69	67	90	66	144	215
Centrica plc	UK	3.03	0.84	6.44	4.64	72	118	226	47	88	98
Smiths Group plc	UK	2.73	4.07	7.63	6.56	19	24	40	57	101	151
BAE Systems plc	UK	1.95	1.12	9.91	10.25	45	29	30	37	320	610
Ashtead Group plc	UK	1.92	0.52	5.73	5.60	71	100	107	76	148	188
Meggitt plc	UK	1.83	3.63	5.92	5.02	31	49	96	27	92	119
Cello Health plc	UK	1.73	0.44	6.37	5.91	83	20	26	45	133	232
Rolls-Royce Holdings plc	UK	1.72	1.91	5.99	5.52	88	34	40	69	90	171
Spirax-Sarco Engineering plc	UK	1.61	-2.39	8.17	7.61	52	133	302	17	123	217
Associated British Foods plc	UK	1.56	1.41	9.03	9.08	53	64	69	28	132	237
Vodafone Group plc	UK	1.36	2.61	7.19	7.54	100	99	168	91	95	119
Compass Group plc	UK	1.14	-0.18	4.60	4.62	50	48	115	33	113	148
Reckitt Benckiser Group plc	UK	0.75	2.91	6.58	6.71	28	30	29	92	148	340
Antofagasta plc	UK	0.70	2.66	4.53	4.73	78	77	66	110	173	310
Schroders plc	UK	0.62	-1.93	3.81	3.90	72	65	91	42	117	160

Table 8.5: UK equity markets trading performance. As before, this table shows the mean return for each strategy for each UK stock selected, expressed as a percentage of the asset price when a trading position was opened. The *predicted long* strategy is the most successful in the majority of cases. An analysis of the maximum gain and loss between each strategy and the *CFD* strategy shows that the *long only* strategy would generally result in more volatile returns. However, the *predicted long* strategy seems to give downside protection whilst maintaining good upside potential.

Table of results – US equities

		CFD			Filtered CFD				95% CI	
Asset	Country	ACCR	TPR	TNR	ACCR.F	TPR.F	TNR.F	P(>ACCR)	ACCR	AUC
General Electric Company	USA	0.62	0.62	0.61	0.52	0.55	0.48	0.00	0.53	0.59
AT&T Inc	USA	0.60	0.64	0.53	0.63	0.70	0.49	0.00	0.53	0.60
Apple Inc	USA	0.60	0.66	0.47	0.52	0.54	0.49	0.00	0.54	0.62
ICICI Bank Limited	USA	0.58	0.67	0.41	0.44	0.46	0.40	0.00	0.54	0.59
Ford Motor Company	USA	0.56	0.63	0.49	0.52	0.49	0.54	0.00	0.53	0.60
Pfizer	USA	0.56	0.63	0.44	0.49	0.59	0.33	0.00	0.53	0.59
The Walt Disney Company	USA	0.55	0.59	0.48	0.62	0.68	0.50	0.00	0.53	0.59
Bank of America Corporation	USA	0.55	0.59	0.48	0.42	0.55	0.20	0.00	0.53	0.59
Advanced Micro Devices	USA	0.53	0.59	0.46	0.57	0.75	0.29	0.12	0.54	0.63
Wells Fargo Company	USA	0.53	0.58	0.42	0.55	0.58	0.50	0.04	0.53	0.58
Oracle corporation	USA	0.52	0.58	0.42	0.49	0.53	0.42	0.14	0.53	0.59
Microsoft Corporation	USA	0.51	0.55	0.42	0.41	0.36	0.52	0.29	0.53	0.59
Intel Corporation	USA	0.49	0.55	0.38	0.48	0.46	0.52	0.70	0.53	0.59
Chesapeake Energy Corporation	USA	0.43	0.48	0.37	0.35	0.40	0.29	1.00	0.54	0.61

Table 8.6: US equity markets prediction performance. These are the results of a selection of individual stocks from the S&P100 index. Applying the algorithm described in §7.3.1 results in only two of the selected stocks satisfying the criteria of both the TPR and TNR being in excess of 50% for the *CFD* strategy. The *filtered CFD* strategy is universally unsuccessful by the same measure.

		Mean return				Maximum Loss (as % of CFD max loss)			Maximum Gain (as % of CFD max gain)		
Asset	Country	CFD	Filtered CFD	Pred. long	Long only	Filtered	Pred. long	Long only	Filtered	Pred. long	Long only
ICICI Bank Limited	USA	7.25	9.26	16.12	11.49	82	90	184	82	101	123
Ford Motor Company	USA	5.05	-0.77	14.10	11.18	92	71	30	100	101	133
Chesapeake Energy Corporation	USA	3.33	-8.65	13.35	11.16	95	52	87	39	52	170
Advanced Micro Devices	USA	3.10	-0.82	11.60	9.96	86	23	50	49	181	335
The Walt Disney Company	USA	2.43	0.38	9.58	8.61	40	84	115	55	136	192
General Electric Company	USA	1.30	-2.66	12.23	13.14	61	125	141	108	134	220
Pfizer	USA	1.27	-0.24	6.91	6.30	69	82	101	89	267	503
Bank of America Corporation	USA	1.02	1.36	6.29	5.84	92	62	86	60	113	168
Apple Inc	USA	0.82	-1.32	7.31	7.65	72	32	42	91	228	526
Oracle corporation	USA	0.15	-0.88	9.49	10.45	56	32	74	82	123	244
AT&T Inc	USA	0.11	-0.41	5.59	6.29	68	71	70	24	131	214
Microsoft Corporation	USA	0.02	-0.95	6.34	6.69	93	25	55	89	237	547
Intel Corporation	USA	-1.86	0.66	5.83	8.51	41	36	28	90	173	320
Wells Fargo Company	USA	-5.60	-3.97	-1.25	4.04	30	68	63	18	100	139

Table 8.7: US equity markets trading performance. The mean return for each strategy for each US stock selected is shown expressed as a percentage of the asset price when a trading position was opened. Although the predictions for this market are not generally successful, the *predicted long* strategy is the most remunerative in half of the cases. In general, all strategies give some downside protection over the *long only* strategy, however, this is at the expense of potential upside.

Table of results – German equities

		CFD			Filtered CFD				95% CI	
Asset	**Country**	**ACCR**	**TPR**	**TNR**	**ACCR.F**	**TPR.F**	**TNR.F**	**P(>ACCR)**	**ACCR**	**AUC**
Adidas AG	Germany	0.67	0.73	0.47	0.58	0.64	0.43	0.00	0.54	0.59
Linde AG	Germany	0.67	0.72	0.48	0.68	0.73	0.48	0.00	0.54	0.59
Henkel AG	Germany	0.62	0.66	0.54	0.61	0.64	0.55	0.00	0.54	0.60
Bayer AG	Germany	0.61	0.67	0.50	0.59	0.65	0.44	0.00	0.54	0.61
Basf SE	Germany	0.61	0.65	0.52	0.59	0.63	0.39	0.00	0.54	0.60
Volkswagen AG	Germany	0.61	0.68	0.44	0.66	0.73	0.42	0.00	0.54	0.60
Deutsche Post AG	Germany	0.59	0.70	0.41	0.56	0.62	0.42	0.00	0.55	0.61
Munchener AG	Germany	0.57	0.62	0.45	0.56	0.58	0.52	0.00	0.54	0.59
BMW AG	Germany	0.55	0.63	0.44	0.55	0.65	0.39	0.02	0.54	0.60
Daimler	Germany	0.55	0.61	0.46	0.55	0.70	0.27	0.02	0.54	0.61
Belesdorf	Germany	0.55	0.62	0.42	0.60	0.67	0.49	0.03	0.54	0.60
Allianz SE	Germany	0.53	0.55	0.47	0.56	0.56	0.55	0.12	0.54	0.60
Deutsche Börse	Germany	0.40	0.52	0.22	0.44	0.49	0.36	1.00	0.55	0.60

Table 8.8: German equity markets prediction performance. These are the results of a selection of individual stocks from the DAX index. Again, by applying the algorithm described in §7.3.1 only two of the selected stocks satisfying the criteria of both the TPR and TNR being in excess of 50% for the *CFD* strategy. However, by using the *filtered CFD* strategy, one can see a very slight increase in in performance.

		Mean return				Maximum Loss (as % of CFD max loss)			Maximum Gain (as % of CFD max gain)		
Asset	**Country**	**CFD**	**Filtered CFD**	**Pred. long**	**Long only**	**Filtered**	**Pred. long**	**Long only**	**Filtered**	**Pred. long**	**Long only**
Linde AG	Germany	10.53	8.26	23.58	16.29	79	85	64	54	144	186
Basf SE	Germany	8.46	4.27	15.51	9.33	68	30	51	54	102	140
Allianz SE	Germany	7.01	4.15	13.18	11.05	100	49	76	76	107	129
Belesdorf	Germany	5.59	0.04	9.20	4.19	65	85	296	24	64	64
Henkel AG	Germany	5.36	3.90	11.37	8.10	69	58	54	82	112	161
Daimler	Germany	4.85	0.76	9.72	5.49	74	53	164	37	86	85
Deutsche Post AG	Germany	3.91	5.42	6.66	3.75	36	123	173	53	93	93
BMW AG	Germany	3.19	3.92	8.73	6.65	87	88	190	70	78	105
Deutsche Börse	Germany	2.98	4.78	3.58	1.55	46	141	206	96	88	83
Volkswagen AG	Germany	2.93	-0.04	8.75	7.08	80	52	73	65	179	248
Bayer AG	Germany	2.89	2.70	7.47	6.21	73	90	130	47	139	209
Munchener AG	Germany	2.79	3.48	9.58	8.51	40	54	205	87	160	286
Adidas AG	Germany	2.73	1.57	9.00	8.11	58	39	52	73	139	232

Table 8.9: German equity markets trading performance. The mean return for each strategy for each German stock selected is shown expressed as a percentage of the asset price when a trading position was opened. Again, the predictions for this market are not generally successful, but in this case the *predicted long* strategy is the most remunerative practically all of the cases. As before, in general, all strategies have less volatile returns than the *long only* strategy.

8.3.3 Commodities

A selection of physical commodities have been considered, and the themes of single stock equities appear to be repeated, in that, although there is a strong performance in terms of TPR, the TNR is poor in all markets, and applying the AUC filter only provides a remedy in three markets. As far as the trading returns are concerned, *predicted long* is a strongly performing strategy providing risk mitigation (although not as strongly as the *filtered CFD* strategy) whilst retaining the upside.

		CFD			Filtered CFD				95% CI	
Asset	**Country**	**ACCR**	**TPR**	**TNR**	**ACCR.F**	**TPR.F**	**TNR.F**	**P(>ACCR)**	**ACCR**	**AUC**
Palladium	NA	0.62	0.75	0.39	0.56	0.66	0.30	0.00	0.55	0.60
Copper	NA	0.61	0.75	0.47	0.61	0.65	0.57	0.00	0.55	0.61
Cotton	NA	0.57	0.68	0.43	0.40	0.52	0.29	0.01	0.55	0.60
Cocoa	NA	0.56	0.67	0.43	0.60	0.78	0.24	0.01	0.54	0.59
Crude Oil	NA	0.54	0.63	0.42	0.49	0.46	0.54	0.07	0.54	0.60
Gold	NA	0.54	0.65	0.35	0.44	0.41	0.47	0.05	0.54	0.59
Sugar	NA	0.53	0.66	0.44	0.63	0.86	0.54	0.18	0.55	0.61
Wheat	NA	0.53	0.78	0.31	0.44	0.72	0.18	0.13	0.54	0.60
Platinum	NA	0.52	0.62	0.41	0.47	0.48	0.46	0.21	0.54	0.59
Soybeans	NA	0.50	0.67	0.34	0.74	0.81	0.59	0.50	0.55	0.60
Heating Oil	NA	0.49	0.88	0.20	0.48	0.83	0.24	0.62	0.56	0.59
Natural Gas	NA	0.49	0.65	0.30	0.68	0.89	0.48	0.62	0.55	0.62
Corn	NA	0.48	0.63	0.32	0.49	0.74	0.22	0.77	0.54	0.59

Table 8.10: Commodity markets prediction performance. The results of a selection of traded commodities are entirely unsuccessful in terms satisfying the criteria of both the TPR and TNR being in excess of 50% for the *CFD* strategy. By using the *filtered CFD* strategy, one can see a very slight increase in performance. In this case, note both the sugar and soybeans markets for their striking performance over the *CFD* strategy.

		Mean return				Maximum Loss (as % of CFD max loss)			Maximum Gain (as % of CFD max gain)		
Asset	**Country**	**CFD**	**Filtered CFD**	**Pred. long**	**Long only**	**Filtered**	**Pred. long**	**Long only**	**Filtered**	**Pred. long**	**Long only**
Palladium	NA	6.16	2.39	10.92	9.14	99	96	104	33	102	126
Copper	NA	5.20	3.10	7.18	3.98	69	126	189	42	77	79
Cocoa	NA	3.31	4.28	5.17	3.15	72	77	137	62	87	85
Natural Gas	NA	2.66	12.12	6.13	5.60	15	90	68	34	100	78
Cotton	NA	2.53	-2.18	4.44	3.04	63	90	183	39	102	147
Corn	NA	2.31	0.78	4.51	3.59	77	94	90	49	100	99
Gold	NA	2.29	0.17	5.25	4.55	74	123	251	49	118	166
Wheat	NA	2.20	-2.97	3.04	2.23	75	101	92	37	100	89
Soybeans	NA	1.85	12.71	3.78	3.18	21	76	76	62	100	110
Crude Oil	NA	1.64	1.88	4.34	3.66	45	72	74	73	102	76
Platinum	NA	1.36	0.34	2.23	1.34	30	98	109	49	100	93
Sugar	NA	1.12	3.78	2.27	1.63	46	82	116	38	103	130
Heating Oil	NA	-0.39	-1.05	-1.14	-1.50	98	127	148	87	112	119

Table 8.11: Commodity markets trading performance. The mean return for each strategy for each commodity is shown expressed as a percentage of the asset price when a trading position was opened. The predictions for this market are very poor, but in this case the *predicted long* strategy is the most remunerative practically all of the cases. The *filtered CFD* strategy is particularly remunerative in the sugar and soybeans markets. As before, in general, all strategies have less volatile returns than the *long only* strategy.

8.3.4 Foreign exchange

The foreign exchange markets are very much the worst performing asset class for the algorithm described in §7.3.1, and with virtually no market demonstrating any interesting predictive qualities, and absolutely no prediction performance enhancement by pursuing the *filtered CFD* strategy. Similarly, there is no consistent theme on which strategy performs the best in terms of the mean percentage return. This could be an indication that foreign exchange markets are not generally susceptible to the development of bubbles and anti-bubbles, and generally do not suffer from trending to the same degree as do the securities and commodity markets.

		CFD			Filtered CFD				95% CI	
Asset	**Country**	**ACCR**	**TPR**	**TNR**	**ACCR.F**	**TPR.F**	**TNR.F**	**P(>ACCR)**	**ACCR**	**AUC**
USD/AUD	NA	0.74	0.81	0.58	0.74	0.84	0.33	0.00	0.55	0.58
USD/EUR	NA	0.60	0.73	0.43	0.56	0.50	0.67	0.00	0.55	0.60
EUR/JPY	NA	0.52	0.75	0.34	0.45	0.86	0.01	0.27	0.55	0.61
EUR/GBP	NA	0.51	0.61	0.40	0.67	0.80	0.42	0.34	0.54	0.60
CHF/JPY	NA	0.51	0.71	0.30	0.66	0.82	0.47	0.37	0.55	0.59
USD/GBP	NA	0.49	0.69	0.25	0.41	0.71	0.12	0.63	0.55	0.59
USD/CAD	NA	0.49	0.65	0.26	0.52	0.69	0.26	0.64	0.55	0.59
AUD/CAD	NA	0.48	0.67	0.30	0.46	0.61	0.34	0.78	0.54	0.58
USD/JPY	NA	0.48	0.60	0.37	0.51	0.84	0.08	0.80	0.54	0.59
CHF/AUD	NA	0.48	0.63	0.29	0.64	0.92	0.02	0.75	0.55	0.59
EUR/CAD	NA	0.46	0.63	0.30	0.59	0.99	0.04	0.92	0.55	0.59
USD/CHF	NA	0.46	0.64	0.27	0.54	0.59	0.49	0.93	0.54	0.58
GBP/CAD	NA	0.45	0.62	0.29	0.58	0.47	0.66	0.97	0.55	0.59
GBP/JPY	NA	0.45	0.62	0.26	0.56	0.91	0.35	0.93	0.56	0.61
EUR/CHF	NA	0.44	0.62	0.34	0.36	0.60	0.22	0.98	0.55	0.59
EUR/AUD	NA	0.44	0.61	0.28	0.40	0.66	0.20	0.96	0.56	0.61
CHF/CAD	NA	0.44	0.59	0.23	0.51	0.72	0.08	0.99	0.54	0.58
GBP/AUD	NA	0.42	0.59	0.26	0.48	0.62	0.39	0.99	0.55	0.60
GBP/CHF	NA	0.39	0.57	0.26	0.50	0.66	0.40	1.00	0.55	0.59

Table 8.12: Foreign exchange markets prediction performance. The results of a selection of currency pairs do not generally satisfy the criteria of both the TPR and TNR being in excess of 50% for the *CFD* strategy. Moreover, given the accuracy of each prediction, one could not be confident that these figures where not a mere artefact of the autocorrelation inherent in the time series. By using the *filtered CFD* strategy, there is no increase in performance.

Asset	Country	Mean return				Maximum Loss (as % of CFD max loss)			Maximum Gain (as % of CFD max gain)		
		CFD	Filtered CFD	Pred. long	Long only	Filtered	Pred. long	Long only	Filtered	Pred. long	Long only
CHF/AUD	NA	3.89	14.33	6.98	5.47	95	100	10	100	100	19
USD/AUD	NA	3.12	3.49	4.42	2.91	93	92	229	65	95	83
USD/EUR	NA	1.66	-0.35	2.38	1.49	92	100	113	13	95	94
GBP/JPY	NA	1.05	2.94	0.65	-0.16	24	100	143	61	71	84
CHF/JPY	NA	1.01	2.92	1.71	1.40	24	68	101	70	94	132
EUR/GBP	NA	0.95	2.51	1.63	1.03	54	143	188	51	94	89
EUR/JPY	NA	0.75	0.42	0.38	-0.21	79	100	87	60	78	63
USD/GBP	NA	0.27	-0.46	1.22	1.48	71	93	86	35	118	162
USD/JPY	NA	0.07	0.60	-0.02	-0.09	79	131	204	83	77	134
GBP/CAD	NA	-0.07	1.95	-0.14	-0.11	56	92	123	94	71	119
USD/CAD	NA	-0.11	0.42	0.54	0.85	41	100	88	58	133	182
AUD/CAD	NA	-0.17	-0.32	-0.06	0.09	46	72	49	26	100	89
CHF/CAD	NA	-0.34	1.38	0.87	1.49	59	52	45	100	112	168
EUR/CHF	NA	-0.63	-1.19	-1.60	-1.43	75	99	108	98	84	53
EUR/AUD	NA	-0.82	-0.43	-0.67	-0.08	59	100	93	97	103	122
EUR/CAD	NA	-0.89	1.08	-0.74	-0.10	46	100	91	85	89	101
USD/CHF	NA	-0.90	0.21	-0.89	-0.32	8	100	76	23	102	88
GBP/CHF	NA	-0.90	0.21	-1.73	-1.43	45	91	101	85	74	67
GBP/AUD	NA	-1.04	0.71	-1.16	-0.51	32	76	72	92	63	128

Table 8.13: Foreign exchange markets trading performance. In contrast to the other markets investigated, the *predicted long* strategy is not the most remunerative. The *filtered CFD* strategy is the best performing strategy of the majority of currency pairs, but this is certainly not reflected in the prediction performance.

Chapter 9

Conclusion and avenues for further research

The aim of this thesis has been to devise methods to detect discrete scale invariance in financial markets. As has been discussed, the signatures of discrete scale invariance can be typified by superexponential growth or decline augmented with log periodic oscillations. This thesis has first explored the current log-periodic power-law models that have been extensively studied since 1996. During the course of this research, a novel approach, derived from the solution to the Schröder equation, has been introduced in §5.2, and resulted in the so called logistic model given in equation (5.19). This model has been particularly successful in modelling asset price anti-bubbles such as the Nikkei 225 index between 1990 and 2004 and the S&P500 asset price bubble preceding the crash of August 1987.

In the later example, the logistic model was shown to provide a compelling solution to the problem of the apparent frequency shifting of the log-periodic oscillations over the decade leading up to the much studied S&P500 Black Monday crash (§5.4.1). By using a model fitting methodology based on visual inspection of the gross features of the time series (§4.3) to give a representative prior distribution of both the angular log-frequency and critical time, it was shown conclusively that the logistic model was able to fit the S&P500 time series in the years prior to the crash very well with a constant angular frequency, whilst retaining a critical time close to that of the crash itself.

This is a very important result in the context of the motivations that inspired the development of the so-called second and third order LPPL models. These models allowed for a frequency shifting in the log-periodic oscillations as the critical time was approached, and were developed to account for a perceived frequency shifting in the Black Monday time-series. Although these models were based on reasonable theoretic grounds, the logistic LPPL model's nonlinear perturbation of the standard linear transformation in the renormalisation group equation, removed the need to consider any frequency shifting at all in this time-series. Further perturbations of the logistic differential equation leads to a hierarchy of logistic LPPL models, and an interesting area of further work would be an investigation of how these model perform in describing asset prices in well known financial market bubbles and anti-bubbles.

Additionally, the logistic model was able to describe the Nikkei 225 index bear market of 1990 – 2004 (§5.4.2) remarkably well when the ex-post predictive performance was compared to that of the second and third-order models derived from the Landau expansion described in §3.3. The deciding factor in this case was the addition of a second singularity in the logistic model. This enabled a very good description of the accelerating drawdowns towards the end of the antibubble to which the saturating LPPL models had no answer.

As mentioned in §5.5 the use of solutions of the renormalisation group formalism corresponding to a nonlinear $\phi(x)$ is certainly appealing, and has provided a model with greater flexibility than the standard LPPL models. However, this thesis has only given a taste of how the logistic model might be used to predict the development of asset price bubbles and anti-bubbles. This model, although of great interest to theoretical practitioners and deserving of much further research, does not appear to be useful in real-time determination of discrete scale invariance, but more suited in analysis of historical data. However, the methods developed Chapters 6, 7 and 8, which aim to indicate whether herding behaviour is governing a financial market, may provide an interesting area for further research in respect identifying suitable historic periods across a wide range of markets which may, or may not, be well described by the logistic model and its higher

extensions.

In order to achieve this kind of real-time asset price bubble or anti-bubble detection, a method was required that is able to analyse the historical data and give a measure of how likely it is that a market is in either a superexponetial or normal growth (or decline) regime at a particular point in time, rather than over at a historical period.

It is this idea that led to the change-point model for asset price bubbles with power-law hazard function described in Chapter 6. In deriving this model, the approach of Cheah & Fry (2015) was taken and a parameter, $\bar{\nu}$, was identified that may go some way to gauging the extent to which a particular market was in, or had been in, a period of superexponential growth, indicating the existence of discrete scale invariance via the control parameter, $\bar{\nu}$. This approach yielded very good visual correlation with areas of high growth in when applied to the S&P500 index between 1960 and the end of 2019. It was noted that there was evidence that there may be a predictive quality to the algorithm developed in this chapter, as it was noticed that regime change was detected prior to crashes, steep sell-offs and longer declines.

It was proposed in Chapter 6 that this parameter could be useful as an early warning signal to the end of a financial bubble, but, as it has been shown, periods of detected discrete scale invariance, if indeed they are financial bubbles, do not necessarily end with a crash at all. A notion that is also shown in (6.32).

Perhaps more useful, from a trading perspective, than such an early warning system, is real-time indication that a financial market is actually in a bubble regime at a particular point in time. Following the efficient market hypothesis, when risk is perceived to be increasing as the market gets deeper into superexponential growth, a higher return should be expected by market participants. This is supported by (6.24) and is the central idea behind the algorithm presented in Chapter 7.

This algorithm stems from the change-point model of Chapter 6 but makes an important jump in terms of its interpretation of the control parameter $\bar{\nu}$. This parameter is described in terms of a linear asset price return growth in a time

dimension transformed by a power-law ("bubble" time) with singularity at the critical time, t_c. The steeper this slope, the greater the superexponential growth in log returns. The flatter the slope, the more likely is the market experiencing normal growth. From this interpretation it has been possible to arrive at an analytical proxy for the control parameter by considering the uniform distributions of the parameter β and the time from the data point in question to the critical time, t_c. This has enabled very swift calculation of the proxy parameter $\bar{\nu}'$ from the historical data, allowing for a large amount of data to be processed in a very short time. This has facilitated an analysis of a broad range of financial markets over long periods of time with only very modest computing power. This is an important aspect of the algorithm, in that one could imagine implementing a platform that would monitor a very large number of financial markets in real-time, seeking out those markets which have entered into regimes of cooperative behaviour.

By using a fixed amount of historical data to calculate the value of $\bar{\nu}'$ at any data point, and normalising these values by reference to a shorter period of historical data, a predictor has been produced. After much research, it has been found that this predictor could be considered useful in predicting a binary outcome of whether the market would be higher or lower at fixed period of time in the future. The results of the trading strategy developed from this underlying method has been presented in Chapter 8. An interesting avenue of further research would be an analysis of the likelihood that a trader would be able to utilise the methodology successfully within financial budget constraints.

These results tell an encouraging story. At least for one asset class examined, the results are remarkably successful. For stock indices it has been shown that accuracy levels over long periods of time are clearly much higher than one would expect to see if the predictor and response were random, autocorrelated, binary sequences. But more importantly the absolute quantum of the accuracy levels would be considered as outstanding for any predictive model of a financial market, and it is obvious that any slight advantage could be exploited by market professionals in order to make huge returns. Hall & Tacon (2010) come to the

not unsurprising conclusion that stock market analysts are not demonstrated to accurately forecast earnings on a persistent basis, whereas the prediction generated by the algorithm in Chapter 7 is shown in Chapter 8 to perform well over very long periods of time.

Although it is true that by employing the *CFD* strategy and back-testing the results over these long periods, even when filtered for the AUC threshold, it has not been shown conclusively to outperform simply buying CFDs on each trading date and closing out this position 130 days later, the *long only* strategy, it is demonstrated that in some cases the volatility of returns is generally smaller, i.e., the lower risk of drawdown is the trade off for a lower the mean percentage return.

This is not the case for the *predicted long* strategy. For stock indices in particular, but also for other assets classes (except foreign exchange), this strategy, in a large number of cases, is the best performing strategy in terms of mean percentage return, and also provides downside protection whilst maintaining strong upside. This is an important result, as it may be an indication that the assumption that an asset price anti-bubble is simply a bubble in reverse is an error in the algorithm construction. Nevertheless, the cross-asset class performance of the *predicted long* strategy has generated tantalising results and gives a very interesting avenue of future research, and a strong hint that when the algorithm predicts a long position should be taken, there is a strong indication that the market has fallen into a regime of discrete scale invariance, signalling the development of an asset price bubble.

It is also interesting to note that the algorithm performs less well for single stocks, even though it is these very stocks that make up some of the stock indices where the algorithm has performed well. This could be a function of the turnover of stocks which make up the index over long periods of time, but it is unlikely that this is the major contributory factor, and as a suggestion for further work, an in-depth analysis of the major index constituents should be undertaken and measured against the index itself.

Although the algorithm has shown some remarkable successes, the algorithm

appears to fall down in respect of the foreign exchange market. But in some ways, maybe this endorses the belief that there is real value in the predictions given and in the algorithm's ability to detect the signatures of discrete scale invariance. Foreign exchange markets are not known to be particularly susceptible to the formation of speculative bubbles (except in the cryptocurrency markets which are not discussed in this thesis) since there are very real economic and international-trade reasons why currencies must be exchanged regardless of market sentiment of some underlying relative value against another currency. This absence of direction could explain the consistently low predication accuracy in these markets.

Therefore, taken in the round, the algorithm described in §7.3.1 has demonstrable merit in respect of its usefulness in detecting periods of discrete scale invariance in financial markets, and this is evidenced by the accuracy of its prediction of market direction. However, there is significant further work that would need to be carried out to confirm this tentative conclusion, and to develop more sophisticated trading strategies that could take advantage of these valuable insights into market mechanics.

Bibliography

Adams, R. P. & MacKay, D. J. C. (2007). Bayesian Online Changepoint Detection, arXiv 0710.3742 [unpublished].

Aminikhanghahi, S. & Cook, D. J. (2017). A survey of methods for time series change point detection, *Knowledge and Information Systems* **51**(2): 339–367.

Anifrani, J.-C., Le Floc'h, C., Sornette, D. & Souillard, B. (1995). Universal log-periodic correction to renormalization group scaling for rupture stress prediction from acoustic emissions, *Journal de Physique I* **5**(6): 631–638.

Arshanapalli, B. & Nelson, W. (2016). Testing for Stock Price Bubbles: a Review of Econometric Tools, *The International Journal of Business and Finance Research* **10**(4): 29–42.

Astill, S., Harvey, D. I., Leybourne, S. J. & Taylor, A. M. (2017). Tests for an end-of-sample bubble in financial time series, *Econometric Reviews* **36**(6-9): 651–666.

Barnsley, M. F. (1988). *Fractals everywhere*, Academic Press.

Barro, R. J., Kormendi, R. C., Kamphuis, R. W. & Watson, J. W. H. (1989). *Black Monday and the future of financial markets / Robert J. Barro ... [et al.] ; edited by Robert W. Kamphuis, Jr., Roger C. Kormendi, and J.W. Henry Watson*, Dow Jones-Irwin ; Mid America Institute for Public Policy Research, Inc Homewood, Ill. : Chicago, Ill.

Benson, A. & Friel, N. (2018). Adaptive MCMC method for multiple changepoint analysis with applications to large datasets, *Electron. J. Statist* **12**(2): 3365–3396.

Bree, D. S., Challet, D. & Peirano, P. P. (2013). Prediction accuracy and sloppiness of log-periodic functions, *Quantitative Finance* **13**(2): 275–280.

Bree, D. S. & Joseph, N. L. (2013). Testing for financial crashes using the log periodic power law model, *International Review of Financial Analysis* **30**: 287–297.

Bufe, C. G. & Varnes, D. J. (1993). Predictive modeling of the seismic cycle of the greater San Francisco Bay region, *Journal of Geophysical Research: Solid Earth* **98**(B6): 9871–9883.

Chang, G. & Feigenbaum, J. A. (2006). A Bayesian analysis of log-periodic precursors to financial crashes, *Quantitative Finance* **6**(1): 15–36.

Cheah, E.-T. & Fry, J. (2015). Speculative bubbles in bitcoin markets? An empirical investigation into the fundamental value of bitcoin, *Economics Letters* **130**: 32 – 36.

Curtright, T. L. & Zachos, C. K. (2011). Renormalization group functional equations, *Physical Review D* **83**(6).

Cvijovic, D. & Klinowski, J. (1995). Taboo search: an approach to the multiple minima problem, *Science* **267**(5198): 664.

Drozdz, S., Ruf, F., Speth, J. & Wójcik, M. (1999). Imprints of log-periodic self-similarity in the stock market, *The European Physical Journal B - Condensed Matter and Complex Systems* **10**(3): 589–593.

Dwyer, G. P. (2015). The economics of bitcoin and similar private digital currencies, *Journal of Financial Stability* **17**: 81 – 91.

Falconer, K. J. (1997). *Techniques in fractal geometry*, Vol. 3, Wiley Chichester (W. Sx.).

Falconer, K. J. (2014). *Fractal Geometry*, Wiley.

Fawcett, T. (2006). An introduction to ROC analysis, *Pattern Recognition Letters* **27**(8): 861 – 874. ROC Analysis in Pattern Recognition.

Feigenbaum, J. A. (2001a). More on a statistical analysis of log-periodic precursors to financial crashes, *Quantitative Finance* **1**(5): 527–532.

Feigenbaum, J. A. (2001b). A statistical analysis of log-periodic precursors to financial crashes, *Quantitative Finance* **1**(3): 346–360.

Feigenbaum, J. A. & Freund, P. G. O. (1996). Discrete scale invariance in stock markets before crashes, *International Journal of Modern Physics B* **10**: 3737–3745.

Feigenbaum, J. A. & Freund, P. G. O. (1998). Discrete scale invariance and the "second Black Monday", *Modern Physics Letters B* **12**(02n03): 57–60.

Filimonov, V. & Sornette, D. (2013). A stable and robust calibration scheme of the log-periodic power law model, *Physica A: Statistical Mechanics and its Applications* **392**(17): 3698–3707.

Gavin, H. (2011). The Levenberg-Marquardt method for nonlinear least squares curve-fitting problems, *Department of Civil and Environmental Engineering* pp. 1–17.

Gellar, M. & Ng, E. W. (1969). A table of integrals of exponential integral, *Journal of Research of the National Bureau of Standards* **73B**(3): 191–.

Geman, H. & Ané, T. (1996). Stochastic subordination: How a "stochastic clock" may account for non-normal returns and stochastic volatility, *RISK Magazine* **9**: 145–149.

Geman, S. & Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images, *IEEE Trans. Pattern Anal. Mach. Intell.* **6**(6): 721–741.

Geraskin, P. & Fantazzini, D. (2013). Everything you always wanted to know about log-periodic power laws for bubble modeling but were afraid to ask, *The European Journal of Finance* **19**(5): 366–391.

Gluzman, S. & Sornette, D. (2002). Log-periodic route to fractal functions, *Phys. Rev. E* **65**: 036142.

Gluzman, S. & Yukalov, V. I. (1998a). Renormalization group analysis of October market crashes, *Modern Physics Letters B* **12**: 75–84.

Gluzman, S. & Yukalov, V. I. (1998b). Resummation methods for analyzing time series, *Modern Physics Letters B* **12**: 61–74.

Gonzalez-Barrios, J. M. (1998). Sums of nonindependant Bernoulli random variable, *Brazilian Journal of Probability and Statistics* **12**(1): 55–64.

Gumbel, E. J. (1944). Ranges and Midranges, *The Annals of Mathematical Statistics* **15**(4): 414–422.

Gürkaynak, R. S. (2008). Econometric tests of asset price bubbles: taking stock, *Journal of Economic Surveys* **22**(1): 166–186.

Hall, J. L. & Tacon, P. B. (2010). Forecast accuracy and stock recommendations, *Journal of Contemporary Accounting and Economics* **6**(1): 18 – 33.

Harvey, D. I., Leybourne, S. J. & Sollis, R. (2015). Recursive Right-Tailed Unit Root Tests for an Explosive Asset Price Bubble, *Journal of Financial Econometrics* **13**(1): 166–187.

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications, *Biometrika* **57**(1): 97–109.

Heard, N. A. & Turcotte, M. J. M. (2017). Adaptive Sequential Monte Carlo for Multiple Changepoint Analysis, *Journal of Computational and Graphical Statistics* **26**(2): 414–423.

Hull, J. C. (2017). *Options, Futures, and Other Derivatives*, Pearson Education Limited.

Hutchinson, J. E. (1981). Fractals and self similarity, *Indiana University Mathematics Journal* **30**(5): 713–747.
URL: *http://www.jstor.org/stable/24893080*

Ilinski, K. (1999). Critical crashes?, *International Journal of Modern Physics C* **10**(04): 741–746.

Ising, E. (1924). *Beitrag zur Theorie des Ferro-und Paramagnetismus*, PhD thesis, Hamburg.

Jackson, B., Scargle, J., Barnes, D., Arabhi, S., Alt, A., Gioumousis, P., Gwin, E., San, P., Tan, L. & Tun Tao Tsai (2005). An algorithm for optimal partitioning of data on an interval, *IEEE Signal Processing Letters* **12**(2): 105–108.

Jarrow, R. (2016). Testing for asset price bubbles: three new approaches, *Quantitative Finance Letters* **4**(1): 4–9.

Johansen, A. (2002). Comment on "Are financial crashes predictable?" by L. Laloux, M. Potters, R. Cont, J.-P. Aguilar and J.-P. Bouchaud, *EPL (Europhysics Letters)* **60**(5): 809.

Johansen, A., Ledoit, O. & Sornette, D. (2000). Crashes as critical points, *International Journal of Theoretical and Applied Finance* **03**(02): 219–255.

Johansen, A. & Sornette, D. (1998). Stock market crashes are outliers, *The European Physical Journal B - Condensed Matter and Complex Systems* **1**(2): 141–143.

Johansen, A. & Sornette, D. (1999a). Critical crashes, *Risk* **12**(1): 91–94.

Johansen, A. & Sornette, D. (1999b). Financial "anti-bubbles": Log-periodicity in gold and Nikkei collapses, *International Journal of Modern Physics C* **10**(04): 563–575.

Johansen, A. & Sornette, D. (1999c). Modeling the stock market prior to large crashes, *The European Physical Journal B - Condensed Matter and Complex Systems* **9**(1): 167–174.

Johansen, A. & Sornette, D. (2000). Evaluation of the quantitative predicition of a trend reversal on the Japanese stock market in 1999, *International Journal of Modern Physics C* **11**(02): 359–364.

Johansen, A., Sornette, D. & Ledoit, O. (1999). Predicting financial crashes using discrete scale invariance, *Risk* **1**(4): 5–32.

Johansen, A., Sornette, D., Wakita, H., Tsunogai, U., Newman, W. I. & Saleur, H. (1996). Discrete scaling in earthquake precursory phenomena: Evidence in the Kobe earthquake, Japan, *Journal de Physique I* **6**(10): 1391–1402.

Jona-Lasinio, G. (1975). The renormalization group: A probabilistic view, *Il Nuovo Cimento B (1971-1996)* **26**(1): 99–119.

Kagan, Y. Y. (1991). Fractal dimension of brittle fracture, *Journal of Nonlinear Science* **1**(1): 1–16.

Killick, R. & Eckley, I. (2013). changepoint: An R Package for changepoint analysis, *Journal of Statistical Software* **58**(3): 1–15.

Killick, R., Fearnhead, P. & Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost, *Journal of the American Statistical Association* **107**(500): 1590–1598.

Laloux, L., M., P., Aguilar, J. P. & Bouchaud, J. P. (2002). Reply to the comment on "are financial crashes predictable?" by A. Johansen, *EPL (Europhysics Letters)* **60**(5): 811.

Laloux, L., Potters, M., Cont, R., Aguilar, J. P. & Bouchaud, J. P. (1998). Are financial crashes predictable?, *EPL (Europhysics Letters)* **45**(1).

Lynch, C. & Mestel, B. (2017). Logistic Model For Stock Market Bubbles And Anti-Bubbles, *International Journal of Theoretical and Applied Finance* **20**(06): 1–24.

Lynch, C. & Mestel, B. (2019). Change-point analysis of asset price bubbles with power-law hazard function, *International Journal of Theoretical and Applied Finance* **22**.

Mandelbrot, B. B. (1983). *The Fractal geomery of nature*, Freeman.

Mandelbrot, B., Fisher, A. & Calvet, L. (1997). A Multifractal Model of Asset Returns, *Cowles Foundation Discussion Papers 1164*, Cowles Foundation for Research in Economics, Yale University.

Matsushita, R., da Silva, S., Figueiredo, A. & Gleria, I. (2006). Log-periodic crashes revisited, *Physica A: Statistical Mechanics and its Applications* **364**: 331–335.

Mills, T. C. & Markellos, R. N. (2008). *The econometric modelling of financial time series*, Cambridge University Press.

Müller, R. (2014). The Boltzmann factor: a simplified derivation, *European Journal of Physics* **35**(5).

Nauenberg, M. (1975). Scaling representation for critical phenomena, *Journal of Physics A: Mathematical and General* **8**(6): 925.

Newman, W. I., Gabrielov, A. M., Durand, T. A., Phoenix, S. L. & Turcotte, D. L. (1993). *An exact renormalization model for earthquakes and material failure: Statics and dynamics*, Conference: Center for Nonlinear Studies international conference on modeling the forces of nature.

Newman, W. I., Turcotte, D. L. & Gabrielov, A. M. (1995). log-periodic behavior of a hierarchical failure model with applications to precursory seismic activation, *Phys. Rev. E* **52**: 4827–4835.

Olson, D. L. & Wu, D. (2013). The impact of distribution on value-at-risk measures, *Mathematical and Computer Modelling* **58**: 1670–1676.

Onsager, L. (1944). Crystal statistics. i. a two-dimensional model with an order-disorder transition, *Phys. Rev.* **65**: 117–149.

Rajesh, J., Jayaraman, V. & Kulkarni, B. (2000). Taboo search algorithm for continuous function optimization, *Chemical Engineering Research and Design* **78**(6): 845–848.

Robertson, D. & Wright, S. M. (1998). The good news and the bad news about long-run stock market returns, *Cambridge Working Papers in Economics* **10.17863/CAM.5034**.

Saleur, H., Sammis, C. G. & Sornette, D. (1996). Discrete scale invariance, complex fractal dimensions, and log-periodic fluctuations in seismicity, *Journal of Geophysical Research: Solid Earth* **101**(B8): 17661–17677.

Smolla, M., B., S., H., L. & Beck, C. (2020). Universal properties of primary and secondary cosmic ray energy spectra, *New Journal of Physics* **22**(9): 093002.

Sornette, D. (2002). Predictability of catastrophic events: Material rupture, earthquakes, turbulence, financial crashes, and human birth, *Proceedings of the National Academy of Sciences of the United States of America* **99**(Suppl 1): 2522–2529.

Sornette, D. (2009). Dragon-kings, black swans and the prediction of crises, *Swiss Finance Institute, Swiss Finance Institute Research Paper Series* **2**.

Sornette, D. (2017). *Why stock markets crash: Critical events in complex financial systems*, Princeton science library.

Sornette, D. & Johansen, A. (1997). Large financial crashes, *Physica A: Statistical Mechanics and its Applications* **245**: 411–422.

Sornette, D. & Johansen, A. (2001). Significance of log-periodic precursors to financial crashes, *Quantitative Finance* **1**(4): 452–471.

Sornette, D., Johansen, A. & Bouchaud, J. (1996). Stock market crashes, precursors and replicas, *Journal de Physique I* **6**(1): 167–175.

Sornette, D. & Sammis, C. G. (1995). Complex critical exponents from renormalization group theory of earthquakes: Implications for earthquake predictions, *Journal de Physique I* **5**(5): 607–619.

Sornette, D., Woodard, R., Yan, W. & Zhou, W. X. (2013). Clarifications to questions and criticisms on the Johansen–Ledoit–Sornette financial bubble model, *Physica A: Statistical Mechanics and its Applications* **392**(19): 4417–4428.

Sornette, D. & Zhou, W. X. (2002). The US 2000-2002 market descent: How much longer and deeper?, *Qunatitative Finance* **2**(6): 468–481.

Taipalus, K. (2012). Detecting asset price bubbles with time-series methods, *Scientific Monographs, Bank of Finland* **E:47**.

Tolikas, K. & Brown, R. A. (2006). The distribution of the extreme daily share returns in the Athens stock exchange, *The European Journal of Finance* **12**(1): 1–22.

Tolikas, K. & Gettinby, G. D. (2009). Modelling the distribution of the extreme share returns in Singapore, *Journal of Empirical Finance* **16**(2): 254–263.

Vandewalle, N., Ausloos, M., Boveroux, P. & Minguet, A. (1998). How the financial crash of October 1997 could have been predicted, *The European Physical Journal B - Condensed Matter and Complex Systems* **4**(2): 139–141.

Vandewalle, N., Ausloos, M., Boveroux, P. & Minguet, A. (1999). Visualizing the log-periodic pattern before crashes, *The European Physical Journal B - Condensed Matter and Complex Systems* **9**(2): 355–359.

Vandewalle, N., Boveroux, P., Minguet, A. & Ausloos, M. (1998). The crash of October 1987 seen as a phase transition: amplitude and universality, *Physica A: Statistical Mechanics and its Applications* **255**(1): 201–210.

Weber, B. (2014). Bitcoin and the legitimacy crisis of money, *Cambridge Journal of Economics* **40**(1): 17–41.

Wilk, G. & Wlodarczyk, Z. (2017). Oscillations in multiparticle production processes, *Entropy* **19**(12).

Wilson, K. G. (1971). Renormalization group and critical phenomena. I. renormalization group and the Kadanoff scaling picture, *Physical Review B* **4**(9): 3174–3183.

Zhou, W. X. & Sornette, D. (2003a). Nonparametric analyses of log-periodic precursors to financial crashes, *International Journal of Modern Physics C* **14**(08): 1107–1125.

Zhou, W. X. & Sornette, D. (2003b). Renormalization group analysis of the 2000–2002 anti-bubble in the US S&P500 index: explanation of the hierarchy of five crashes and prediction, *Physica A: Statistical Mechanics and its Applications* **330**(3–4): 584–604.

Zhou, W. X. & Sornette, D. (2005). Testing the stability of the 2000 US stock market antibubble, *Physica A: Statistical Mechanics and its Applications* **348**: 428–452.

Zhou, W. X. & Sornette, D. (2006). Fundamental factors versus herding in the 2000–2005 US stock market and prediction, *Physica A: Statistical Mechanics and its Applications* **360**(2): 459–482.